



**PHD**

**Novel thermostable DNA Polymerases for isothermal DNA amplification**

Morant, Nick

*Award date:*  
2015

*Awarding institution:*  
University of Bath

[Link to publication](#)

**Alternative formats**

If you require this document in an alternative format, please contact:  
[openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk)

Copyright of this thesis rests with the author. Access is subject to the above licence, if given. If no licence is specified above, original content in this thesis is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International (CC BY-NC-ND 4.0) Licence (<https://creativecommons.org/licenses/by-nc-nd/4.0/>). Any third-party copyright material present remains the property of its respective owner(s) and is licensed under its existing terms.

**Take down policy**

If you consider content within Bath's Research Portal to be in breach of UK law, please contact: [openaccess@bath.ac.uk](mailto:openaccess@bath.ac.uk) with the details. Your claim will be investigated and, where appropriate, the item will be removed from public view as soon as possible.

# **Novel thermostable DNA polymerases for isothermal DNA amplification**

Submitted by  
Nicholas Richard Morant

For the degree of Doctor of Philosophy  
University of Bath  
Department of Biology and Biochemistry  
October 2014

## **COPYRIGHT**

Attention is drawn to the fact that copyright of this thesis rests with the author. A copy of this thesis has been supplied on condition that anyone who consults it is understood to recognise that its copyright rests with the author and that they must not copy it or use material from it except as permitted by law or with the consent of the author.

This thesis may be made available for consultation within the University Library and may be photocopied or lent to other libraries for the purpose of consultation with effect from

# Acknowledgments

I would like to thank my supervisors Professor Michael Danson and Dr Duncan Clark; I am extremely grateful for their support and encouragement and for the opportunities they have given me throughout my Ph.D.

I would also like to thank my colleagues at GeneSys Biotech Ltd. for their assistance and for the generosity with which they share their expertise. In particular, I would like to thank Sarah Dugdale, Anna Siddle and Dr Natasha Hilton for their advice and support during the preparation of this thesis. Special thanks go to Dr Duncan Clark for his continued enthusiasm and motivation at GeneSys Biotech Ltd.

I am extremely grateful to have such wonderful family and friends who have encouraged me throughout. A special thanks to my Mum and Dad for their love and support over the years.

On a personal note, thank you Mared, for getting me through, and for discovering an enzyme all of your own...the elusive Poly'A'merase.

# CONTENTS

<b>Abstract.....</b>	<b>6</b>
<b>Abbreviations.....</b>	<b>7</b>
<b>1. INTRODUCTION</b>	
<b>1.1 DNA Polymerases.....</b>	<b>9</b>
1.1.1 History, structure and function.....	9
1.1.2 Nucleotide incorporation.....	18
1.1.3 Strand-displacement activity.....	23
<b>1.2 DNA amplification methods.....</b>	<b>27</b>
1.2.1 Nucleic acid detection.....	27
1.2.2 Isothermal nucleic acid amplification.....	28
1.2.3 Loop-mediated isothermal amplification.....	30
1.2.4 Applications for biotechnology.....	35
<b>1.3 Project objectives.....</b>	<b>36</b>
<b>2. MATERIALS AND METHODS</b>	
<b>2.1 Materials.....</b>	<b>37</b>
2.1.1 Media.....	38
2.1.2 Buffers.....	38
<b>2.2 Methods: Identification.....</b>	<b>39</b>
2.2.1 Strain isolation.....	39
2.2.2 Genomic DNA extraction.....	39
2.2.3 16S rRNA gene amplification.....	40
2.2.4 <i>polA</i> sequence identification.....	40
2.2.5 TA Cloning.....	40
2.2.6 Ethanol precipitation.....	41
2.2.7 Transformation.....	42
2.2.8 Preparation of electrocompetent <i>E.coli</i> .....	42
2.2.9 Colony screening.....	43
2.2.10 DNA sequencing.....	43
2.2.11 Gene-walking.....	44
2.2.12 Sequence alignments.....	45
<b>2.3 Methods: Molecular cloning.....</b>	<b>46</b>
2.3.1 Directional cloning primer design.....	47
2.3.2 High Fidelity Cloning PCR.....	47
2.3.3 Restriction endonuclease digest.....	47
2.3.4 Ligations.....	48
2.3.5 Plasmid mini-prep.....	49
2.3.6 Site-directed mutagenesis.....	49
2.3.7 -70°C Freezer stocks.....	50



<b>2.4</b>	<b>Methods: Protein expression.....</b>	<b>50</b>
2.4.1	Small-scale protein expression.....	50
2.4.2	Large scale protein expression.....	51
2.4.3	96-well plate protein expression.....	51
2.4.4	Protein sample preparation.....	51
2.4.5	SDS Polyacrylamide Gel Electrophoresis.....	52
<b>2.5</b>	<b>Methods: Protein Purification.....</b>	<b>52</b>
2.5.1	Metal-chelating affinity column chromatography.....	52
2.5.2	Heparin column chromatography.....	53
2.5.3	Ion exchange column chromatography.....	54
2.5.4	Protein concentration.....	54
<b>2.6</b>	<b>Methods: Characterisation.....</b>	<b>55</b>
2.6.1	Nuclease assay.....	55
2.6.2	Genomic DNA contamination assay.....	55
2.6.3	Thermal shift assay.....	55
2.6.4	Strand-displacing primer extension assay.....	56
2.6.5	Loop-mediated Isothermal DNA amplification.....	56
2.6.6	Reverse transcription LAMP.....	57
2.6.7	Shimadzu analyser of amplification reactions.....	57
2.6.8	Buffer optimisation.....	57
2.6.9	Incorporation of dUTP.....	57
2.6.10	<i>B.st</i> LF DNA Polymerase I unit assay.....	58
2.6.11	Polymerase chain reaction (PCR).....	58
<b>3.</b>	<b>IDENTIFICATION</b>	
<b>3.1</b>	<b>Introduction.....</b>	<b>59</b>
3.1.1	Chapter overview.....	62
<b>3.2</b>	<b>Results.....</b>	<b>63</b>
3.2.1	Isolation of thermophilic strains.....	63
3.2.2	Isolation of genomic DNA.....	67
3.2.3	16S rRNA gene sequence analysis.....	68
3.2.4	<i>polA</i> identification.....	69
3.2.5	<i>polA</i> gene-walking.....	70
3.2.6	<i>G.me</i> LF DNA polymerase I.....	72
3.2.7	<i>T.at</i> LF DNA polymerase I.....	76
3.2.8	Additional LF DNA polymerase I.....	78
3.2.9	Sequenced strains <i>polA</i> identification.....	79
3.2.10	<i>polA</i> sequence alignments.....	80
<b>3.3</b>	<b>Discussion.....</b>	<b>88</b>
<b>4.</b>	<b>CLONING, EXPRESSION &amp; PURIFICATION</b>	
<b>4.1</b>	<b>Introduction.....</b>	<b>91</b>
4.1.1	Polymerase genes identified.....	91
4.1.2	Recombinant cloning and protein expression.....	92
4.1.3	Purification.....	93
4.1.4	Protein quantification.....	95

4.1.5	Chapter overview.....	95
<b>4.2</b>	<b>Results.....</b>	<b>96</b>
4.2.1	Cloning of <i>G.me</i> and <i>T.at</i> LF DNA polymerase I.....	96
4.2.2	Small-scale expression of <i>G.me</i> and <i>T.at</i> LF DNA polymerase I.....	98
4.2.3	Remaining LF DNA polymerase I cloning and expression.....	99
<b>4.3</b>	<b>Purification.....</b>	<b>101</b>
4.3.1	Purification of <i>G.me</i> and <i>T.at</i> LF DNA polymerase I.....	101
4.3.2	Purification of remaining LF DNA polymerase I.....	110
4.3.3	Protein quantification.....	111
<b>4.4</b>	<b>Discussion.....</b>	<b>113</b>
<b>5.</b>	<b>CHARACTERISATION</b>	
<b>5.1</b>	<b>Introduction.....</b>	<b>115</b>
5.1.1	Polymerase unit concentrations.....	115
5.1.2	Protein thermostability.....	115
5.1.3	Strand-displacement activity.....	118
5.1.4	Chapter overview.....	119
<b>5.2</b>	<b>Results.....</b>	<b>120</b>
5.2.1	Nuclease contamination.....	120
5.2.2	DNA contamination.....	121
5.2.3	Protein thermal melt.....	124
5.2.4	Strand-displacement activity.....	128
5.2.5	Isothermal DNA amplification.....	135
5.2.6	Heat-denaturing LAMP.....	149
5.2.7	Buffer optimisation.....	151
5.2.8	Incorporation of dUTP.....	154
5.2.9	Reverse transcription LAMP.....	155
5.2.10	Polymerase unit concentrations.....	157
5.2.11	Protein crystallisation.....	159
<b>5.3</b>	<b>Discussion.....</b>	<b>160</b>
<b>6.</b>	<b>ENGINEERING</b>	
<b>6.1</b>	<b>Introduction.....</b>	<b>171</b>
6.1.1	The Singh SYR strand-displacement motif.....	171
6.1.2	Molecular mutagenesis techniques.....	172
6.1.3	Chapter overview.....	174
<b>6.2</b>	<b>Results.....</b>	<b>175</b>
6.2.1	<i>G.me</i> LF DNA polymerase I engineering.....	175
6.2.2	3'-5' exonuclease mutants.....	180
6.2.3	Strand-displacing <i>T.aq</i> and <i>T.ma</i> LF DNA polymerase I.....	193
6.2.4	Strand displacement 'SYR' motif.....	200
<b>6.3</b>	<b>Discussion.....</b>	<b>208</b>

<b>7.</b>	<b>CHIMERAS</b>	
<b>7.1</b>	<b>Introduction.....</b>	<b>213</b>
7.1.1	Chapter overview.....	214
<b>7.2</b>	<b>Results.....</b>	<b>215</b>
7.2.1	<i>T.in(exo-)/G.me*</i> LF DNA polymerase I chimeras.....	216
7.2.2	<i>T.in-T.aq</i> LF DNA polymerase I chimeras.....	223
<b>7.3</b>	<b>Discussion.....</b>	<b>233</b>
<b>8.</b>	<b>FUSIONS</b>	
<b>8.1</b>	<b>Introduction.....</b>	<b>238</b>
8.1.1	Chapter overview.....	239
<b>8.2</b>	<b>Results.....</b>	<b>240</b>
8.2.1	Cloning, expression and purification.....	240
8.2.2	Characterisation.....	243
<b>8.3</b>	<b>Discussion.....</b>	<b>249</b>
<b>9.</b>	<b>FINAL DISCUSSION</b>	
<b>9.1</b>	<b>Report summary.....</b>	<b>252</b>
<b>9.2</b>	<b>Industrial applications.....</b>	<b>257</b>
<b>9.3</b>	<b>Future aspects.....</b>	<b>261</b>
<b>IV</b>	<b>References.....</b>	<b>264</b>
<b>V</b>	<b>Appendix.....</b>	<b>280</b>
i	Primer sequences.....	280
ii	DNA Polymerase I sequences.....	285
iii	DNA polymerase I chimera sequences.....	293
iv	DNA polymerase I fusion sequences.....	300
v	Plasmids.....	301
vi	Set for Britain Award 2013.....	302

## Abstract

DNA polymerases play a fundamental role in the transmission and maintenance of genetic information and have become an important *in vitro* diagnostic and analytical tool. The Loop-mediated isothermal DNA amplification (LAMP) method has major applications for disease and pathogen detection and utilises the unique strand-displacement activity of a small group of thermostable DNA polymerases. The Large (Klenow-like) Fragment of *Geobacillus stearothermophilus* DNA polymerase I (*B.st* LF Pol I) currently serves as the enzyme of choice for the majority of these isothermal reactions, with few alternatives commercially available. An increasing need for point-of-care nucleic acid diagnostics is now shifting detection methods away from traditional laboratory based chemistries, such as the polymerase chain reaction (PCR), in favour of faster, and often simpler, isothermal methods. It was recognised that in order to facilitate these rapid isothermal reactions there was a requirement for alternative thermostable, strand-displacing DNA polymerases and this was the basis of this thesis.

This thesis reports the successful identification of polymerases from Family A, chosen for their inherent strand-displacement activity, which is essential for the removal of RNA primers of Okazaki fragments during lagging-strand DNA synthesis *in vivo*. Twelve thermophilic organisms, with growth temperature ranges between 50°C and 80°C, were identified and the genomic DNA extracted. Where DNA sequences were unavailable, a gene-walking technique revealed the *polA* sequences, enabling the Large Fragment Pol I to be cloned and the recombinant protein over-expressed in *Escherichia coli*. A three-stage column chromatography purification permitted the characterisation of ten newly identified Pol I enzymes suitable for use in LAMP. *Thermodesulfatator indicus* (*T.in*) Pol I proved to be the most interesting enzyme isolated. Demonstrating strong strand-displacement activity and thermostability to 98°C, *T.in* Pol I is uniquely suitable to a newly termed heat-denaturing LAMP (HD-LAMP) reaction offering many potential advantages over the existing LAMP protocol.

The current understanding of strand-displacement activity of Pol I is poorly understood. This thesis recognised the need to identify the exact regions and motifs responsible for this activity of the enzyme, enabling potential enhancements to be made. Enzyme engineering using site-directed mutagenesis and the formation of chimeras confirmed the importance of specific subdomains in strand-separation activity. With this knowledge, a unique *Thermus aquaticus* (*T.aq*) Pol I mutant demonstrated sufficient strand-displacement activity to permit its use in LAMP for the first time. The fusion of Cren7, a double-stranded DNA binding protein, to Pol I for use in LAMP is also reported. Although the fusion construct was found to reduce amplification speed, enhancements were observed in the presence of increased salt concentrations and it is suggested here as a means for future enzyme development.

## Abbreviations

AA	Amino-acid
AMV-RT	Avian myeloblastosis virus reverse transcriptase
APS	Ammonium persulphate
bp	Base pairs (DNA)
BSA	Bovine serum albumin
cDNA	Complementary DNA
CTAB	Cetyltrimethylammonium bromide
CV	Column volume
<i>Cr</i>	Chloramphenicol
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded DNA
DSMZ	Deutsche Sammlung von Mikroorganismen und Zellkulturen
dNTP	Deoxyribonucleic acid triphosphate
DMSO	Dimethyl sulfoxide
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
Exo	Exonuclease
FL	Full length
gDNA	Genomic DNA
HDA	Helicase-dependent amplification
HD-LAMP	Heat-denaturing loop-mediated isothermal amplification
HF	High fidelity
HNB	Hydroxynaphthol blue
IGEPAL-CA630	Octylphenoxy poly(ethyleneoxy)ethanol
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
<i>Kan</i>	Kanamycin sulphate
kDa	KiloDalton
KO	Knock-out
LAMP	Loop-mediated isothermal amplification
LB	Luria Broth
LF	Large fragment
Mw	Molecular weight
NA	Nucleic acid
NASBA	Nucleic acid sequence-based amplification
NEAR	Nicking and extension amplification reaction

NEC	No enzyme control
nm	Nanometre
NTC	No template control
OD	Optical density
ORF	Open reading frame
pI	Isoelectric point
POCT	Point-of-care test(ing)
Pol	Polymerase
PCR	Polymerase chain reaction
pDNA	Plasmid DNA
qPCR	Quantitative PCR
rDNA/rRNA	Ribosomal DNA/Ribosomal RNA
RCA	Rolling circle amplification
RE	Restriction endonucleases
RNA	Ribonucleic acid
RPA	Recombinase polymerase amplification
RT-LAMP	Reverse transcription loop-mediated isothermal amplification
RT-PCR	Reverse transcription polymerase chain reaction
SAP	Shrimp alkaline phosphatase
SD	Strand-displacement
SDA	Strand-displacement amplification
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNP	Single nucleotide polymorphism
SSB	Single-stranded DNA-binding protein
ssDNA	Single-stranded DNA
T <sub>A</sub>	Temperature of annealing
T <sub>m</sub>	Temperature of melting
TB	Terrific broth
TEMED	N,N,N',N'-Tetramethylethylenediamine
tRNA	Transfer RNA
X-Gal	5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside

# 1. INTRODUCTION

## 1.1 DNA Polymerases

### 1.1.1 History, structure and function

DNA polymerases play a fundamental role in the transmission and maintenance of genetic information. They catalyse the template-directed addition of deoxynucleotides onto a DNA primer and they function in replication, repair, and recombination of DNA (Kornberg *et al.*, 1958, Kornberg, 1980). *Escherichia coli* (*E.co*) DNA polymerase I, the first polymerase identified, was discovered just 3 years after Watson and Crick established the structure of DNA. They concluded their ground breaking paper with “It has not escaped our notice that the specific pairing we have postulated immediately suggests a possible copying mechanism for the genetic material” (Watson and Crick, 1953).

On the basis of amino acid similarity and crystal structure analysis, DNA polymerases have been grouped into seven different families, namely the DNA-dependant DNA polymerases A, B, C, D, X, Y and the RNA-dependant DNA polymerase reverse-transcriptases (Delarue *et al.* 1990; Ito and Braithwaite, 1991; Braithwaite and Ito, 1993; Joyce and Steitz, 1994) (Table 1.1). New polymerases continue to be discovered and occasionally their sequence-based placement into one of the families proves difficult.

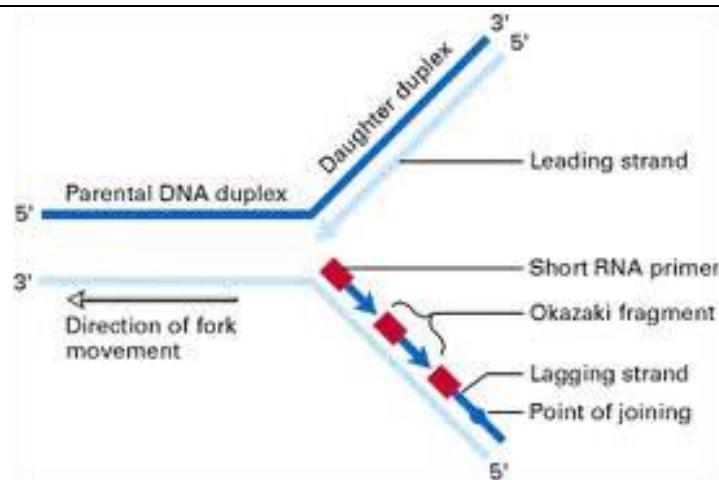
Family	Prokaryotic	Eukaryotic	Archaea	Viral
A	Pol I	Pol $\gamma$ , $\theta$		T3, T5, T7 pol
B	Pol II	$\alpha$ , $\delta$ , $\epsilon$ , $\xi$	Pol BI, BII	$\phi$ 29, T4, T6 pol
C	Pol III			
D			Pol D	
X		$\beta$ , $\lambda$ , $\sigma$ , $\mu$ . TdT		
Y	Pol IV, V	Pol $\eta$ , $\iota$ , $\kappa$	Dpo4	
RT		Telomerase		Reverse transcriptases

**Table 1.1**

Representative members of families of DNA polymerase. Adapted from Patel *et al.* 2001.

### Family A

The Family A DNA polymerases (Pol I) are encoded by the *polA* gene, and can be grouped into recombination, replicative and repair enzymes (Kornberg, 1980). Examples include: Bacteriophage T5 and T7 Pol I, *E.co* Pol I, *Thermus aquaticus* (*T.aq*) Pol I, and *Geobacillus stearothermophilus* (*B.st*) Pol I. They are involved in nucleotide excision repair and in processing Okazaki fragments that are generated during lagging strand synthesis (Kornberg and Baker, 1992) (Figure 1.1). The majority of DNA Pol I enzymes contain 5'-3' exonuclease activity, 3'-5' exonuclease (proofreading) activity and have a simple, single subunit structure.



**Figure 1.1**

Leading and lagging strand DNA synthesis. The polymerisation activity of the replicative polymerase (e.g. Pol III) takes place only in the 5'-3' direction. Therefore one of the two strands of DNA having 3'-5' polarity gets continuous synthesis of DNA whilst the other strand, having 5'-3' polarity, gets synthesis of DNA in small fragments called Okazaki fragments. The synthesis of DNA on this strand is opposite to the movement of the replication fork and is called the lagging strand, while the continuous strand is called the leading strand. DNA Pol I acts to remove the RNA primers and fill in the gaps between the Okazaki fragments ensuring both DNA strands are faithfully replicated.

### Family B

The Family B DNA polymerases (Pol II), encoded by the *polB* gene, are predominantly involved in DNA replication and repair. The enzyme participates in leading and lagging strand synthesis (Kornberg, 1980). Pol II enzymes have DNA repair roles and are usually single-subunit enzymes. Examples include *E.co* Pol II, archaeal polymerases such as those from *Pyrococcus furiosus* (*P.fu*), and viral polymerases such as Bacteriophage Phi29 and T4 Pol II. The Family B DNA polymerases contain a strong 3'-5' exonuclease activity that



corrects errors during DNA replication. The 3'-5' exonuclease activity has been shown to be over 1000 times higher than that of *E.co* DNA pol I (Capson *et al.*, 1992. Lin *et al.*, 1994).

### **Family C**

Family C polymerases (Pol III) are the major chromosomal replicative enzymes in prokaryotes e.g. *E.co* Pol III. The holoenzyme interacts with other proteins and forms a large multi-subunit complex consisting of at least 10 subunits (Kornberg, 1992).

### **Family D**

Family D polymerases have not been well characterised but are found in the euryarchaeota subdomain of Archaea. DNA Pol II of Family D has been suggested to be a multi-subunit replicative polymerase (Uemori *et al.*, 1997).

### **Family X**

Family X DNA Polymerases contain eukaryotic DNA polymerase  $\beta$ , a eukaryotic repair enzyme that has no corresponding *E.co* DNA polymerase (Braithwaite and Ito, 1993), polymerase  $\sigma$ , polymerase  $\mu$ , polymerase  $\lambda$ , yeast polymerase, and the African swine fever virus polymerase X (Rothwell *et al.*, 1995). Although its primary amino acid sequence places it in a unique Family, Pol  $\beta$  can carry out several Pol I like functions including gap-filling between Okazaki fragments and excision repair (Sweasy *et al.*, 1992).

### **Family Y**

Family Y DNA polymerases (DNA Pol IV and V) are able to recognise and bypass different classes of lesions in DNA (Friedberg *et al.*, 2000). Members are found in eubacteria, eukaryotes and archaea. Family Y enzymes characteristically show low fidelity and specificity in order to deal with damaged DNA templates and, as such, lack 3'-5' exonuclease activity. The enzymes are thought to function in a distributive manner so as not to cause mutagenic incorporation events after lesion bypass is completed. Examples include *E.co* Pol IV, encoded by *dinB* (Steitz *et al.*, 2001), *Sulfolobus solfataricus* (*S.so*) Dpo4 (Woodgate *et.al*, 2001) and *E.co* Pol V, encoded by the *UmuCD* genes (Goodman, 1999).

### **Family RT**

The reverse transcriptase (RT) Family includes RTs from retroviruses as well as eukaryotic telomerases. They act to convert a single-stranded RNA template into double-stranded DNA. Examples include Human Immunodeficiency Virus (HIV-1 and 2 RT) and Moloney Murine Leukemia Virus (MMuLV-RT). Both contain a polymerase domain as well as an

RNAseH domain to cleave viral RNA during DNA synthesis (Rothwell *et al.*, 2005).

## Structure

The comparison of tertiary structures of DNA polymerases suggests a greater overall similarity than purely their primary amino acid sequences. While primary sequences and function have diverged, structural similarities remain (Hamilton *et al.*, 2001). The most extensively studied polymerases include those in Family A (found in prokaryotes, eukaryotes and bacteriophages), and those in Family B (found in prokaryotes, eukaryotes, archaea and viruses). All DNA polymerases, however, share an overall morphology. Early crystallographic studies described the structure of DNA Pol I as resembling a 'right hand holding a rod with a very deep cleft (Palm), curled-over Fingers and flexible Thumb' (Ollis *et al.*, 1985) (Figure 1.2). The tips of the Fingers and Thumb subdomains are in close contact so that the cleft can be described as a tunnel (Joyce *et al.*, 1994). Comparisons of a wide variety of polymerase structures identified these Palm, Fingers and Thumb subdomains to serve analogous functions due to their close structural homology, despite their lack of a shared evolutionary origin (Joyce and Steitz, 1995).

### *Palm subdomain*

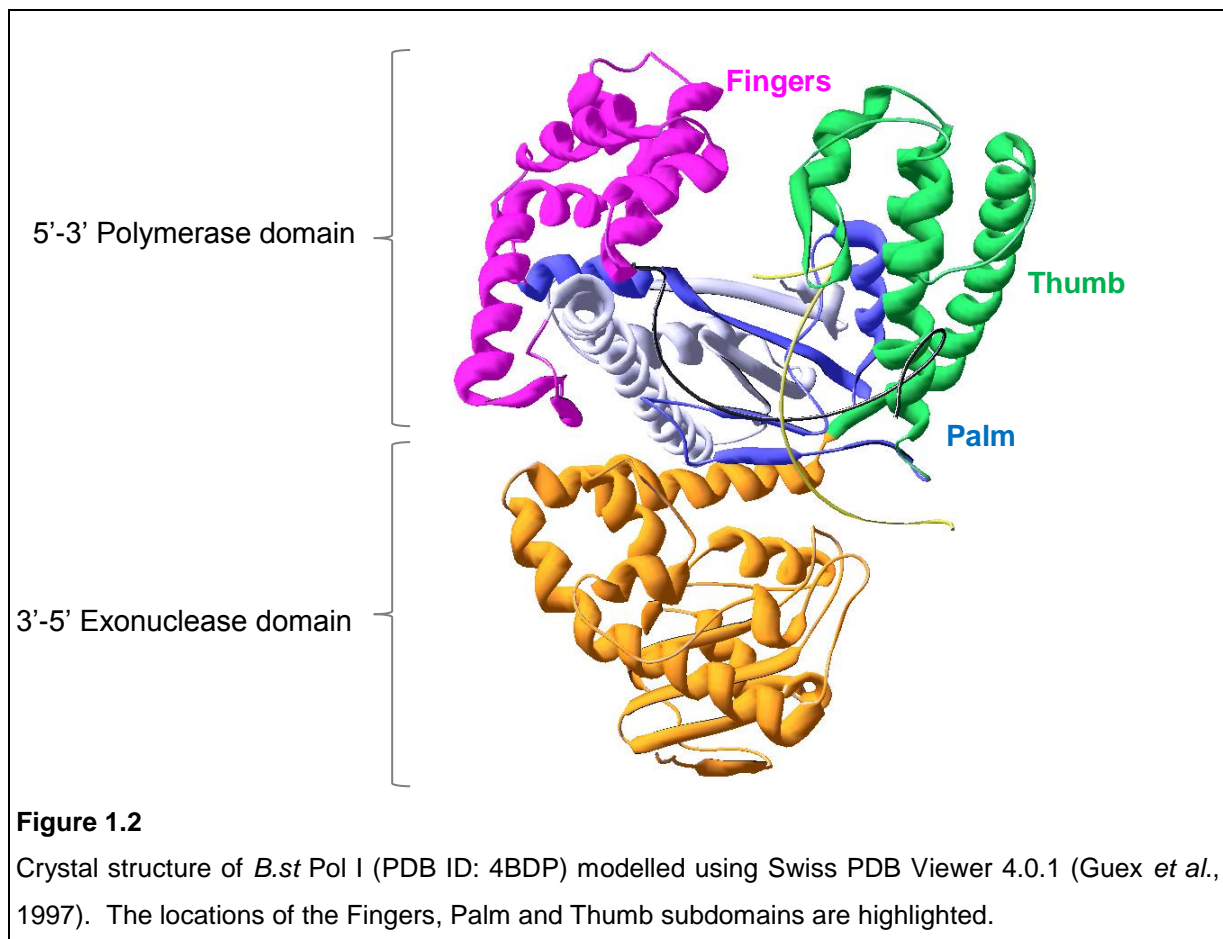
The Palm subdomain is the most highly conserved domain between polymerases and forms the base of the polymerase cleft. The domain is the catalytic portion of the polymerase. It contains three highly-conserved catalytic residues (Asp<sup>705</sup>, Asp<sup>882</sup> and Glu<sup>883</sup> of *E.co* Pol I) that bind two divalent ions (Mg<sup>2+</sup>) that play critical roles in the catalysis step (Beese *et al.*, 1993). The shape of the catalytic active site of DNA polymerases has been shown to be superimposable despite an often low level of primary sequence similarity (reviewed by Joyce and Steitz, 1995). Minor changes in the amino acid sequence within the Palm subdomain have been shown to dramatically affect substrate binding specificity (Tabor *et al.*, 1995; Astatke *et al.*, 1998; Patel *et al.*, 2000).

### *Thumb subdomain*

The less-conserved Thumb subdomain interacts with the minor groove of DNA and with the incoming nucleotide of the template-primer upstream of the site of synthesis, and is composed largely of helical structures (Joyce *et al.*, 1994). Loops at the tip of the Thumb also interact with the DNA backbone and direct partitioning between polymerisation and editing (3'-5' exonuclease) modes (Steitz *et al.*, 1995).

### *Fingers subdomain*

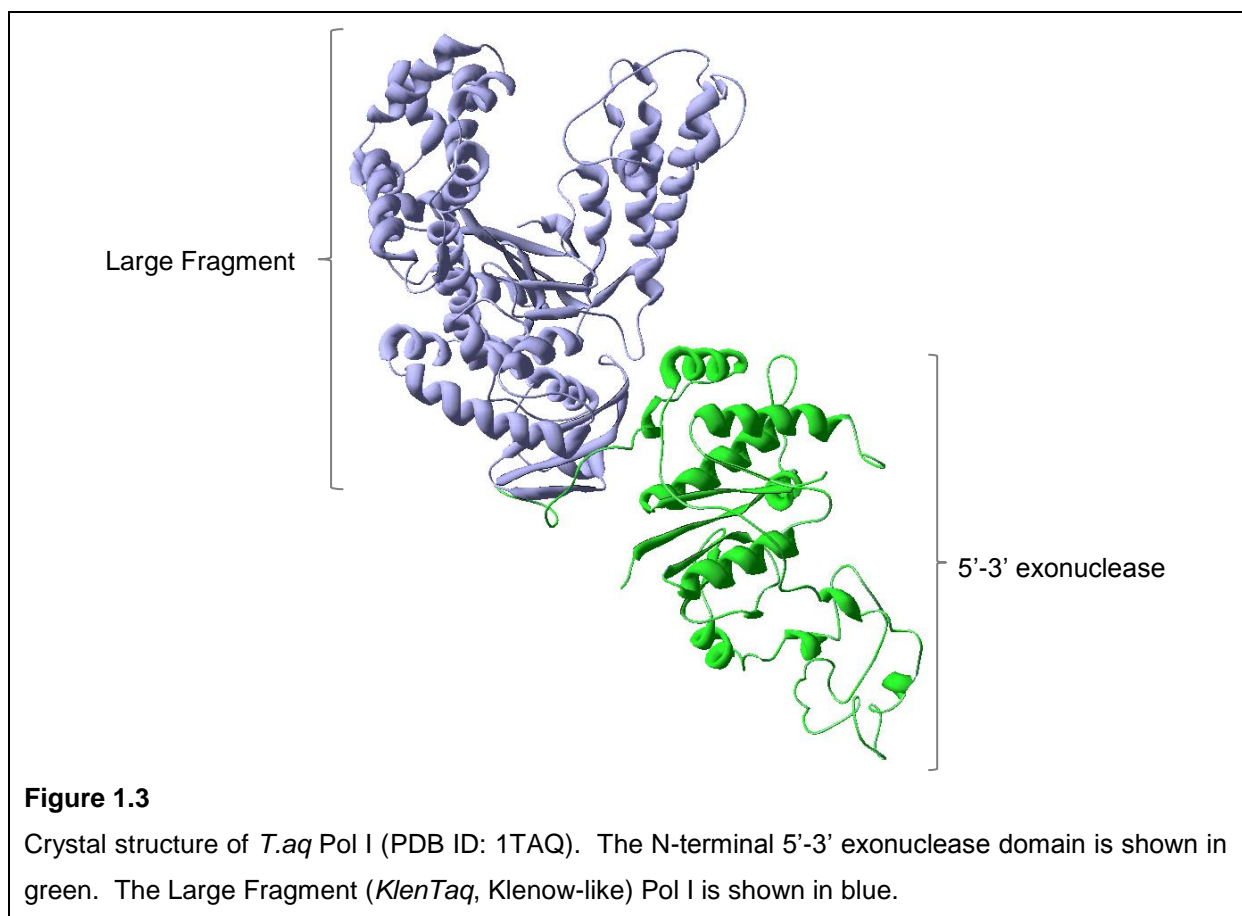
The Fingers subdomain binds the single-stranded template across from and beyond the site of synthesis. It forms one side of the pocket surrounding the nascent base pair, and is believed to play a role in template fixation, template specificity (Delarue *et al.*, 1990), and the binding and orientation of the template strand (Joyce and Steitz, 1995). Whilst showing the greatest structural diversity of the three subdomains, the Fingers subdomain contains the highly conserved Motif A and Motif C that are conserved among DNA-dependent RNA and DNA polymerases as well as RNA-dependent DNA and RNA polymerases (Delarue *et al.*, 1990). Joyce *et al.* (1994) suggest the Fingers subdomain has a universal role in template binding but is structurally adapted to the specific template involved.



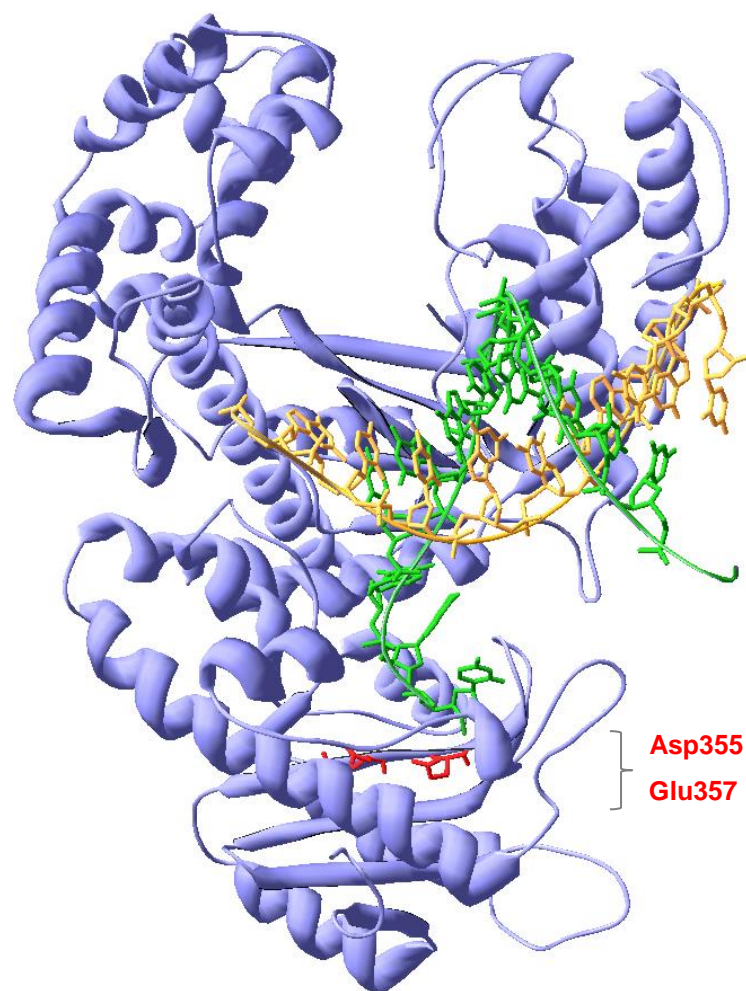
Crystal structures have shown an additional subdomain to be present in Family Y polymerases (Steitz *et al.*, 2001). This C-terminal 'Little Finger' or 'Wrist' is tethered to the Thumb but is physically located next to the Fingers subdomain. Unusually-small Fingers and Thumb subdomains lead to a more open active site than Family A or B polymerases, permitting reduced fidelity of nucleotide incorporation (Woodgate *et al.*, 2001).

### Exonuclease domain

A typical DNA Polymerase I enzyme contains either an N-terminal domain with 5'-3' exonuclease activity or with 3'-5' exonuclease activity, or both. The 5'-3' exonuclease domain removes nucleotides ahead of the extending polymerase during lagging strand synthesis of replication or when removing RNA primers from Okazaki fragments (Kornberg, 1980). In the bacterial DNA Pol I enzymes, this activity is part of the same polypeptide chain as the polymerase (Joyce *et al.*, 1994), whereas bacteriophage T4, T5, and T7 encode the same activity in a separate enzyme (Hollingsworth *et al.*, 1991). The 5'-3' exonuclease is often described as a 'structure-specific' endonuclease because the preferred substrate is a displaced 5' end of DNA in a flap orientation generated by strand-displacement synthesis (Lundquist *et al.*, 1982). Cleavage takes place at or close to the junction between the single-stranded and duplex DNA, requiring a free 5' end for access to the substrate (Joyce *et al.*, 1994). In 1970, Klenow *et al.* discovered the 109kDa *E.co* DNA Pol I protein could be reduced by the protease subtilisin to yield two fragments: a Large Fragment (LF), retaining the 5'-3' polymerase and 3'-5' exonuclease activity, and a smaller fragment containing the 5'-3' exonuclease domain (Figure 1.3).



The 3'-5' exonuclease, or proofreading, activity allows greater fidelity of nucleotide incorporation. This is achieved by removing incorrectly incorporated nucleotides from the growing DNA primer strand resulting in another opportunity to incorporate the correct base, before continuing on with synthesis (Joyce *et al.*, 1994). Of the polymerases that contain this function, the 3'-5' exonuclease active site is highly conserved (Ishino *et al.*, 1994; Derbyshire *et al.*, 1995). The proofreading domain of many polymerase Family A members, including those from *T.aq* and *B.st*, is present but inactive due to the lack of key catalytic residues (Derbyshire *et al.*, 1988). Conserved regions that are immediately upstream from the polymerase domain encode the 3'-5' exonuclease activity (Figure 1.4). Domain removal cannot be achieved because large deletions from this region of the enzyme have been shown to cause inactivation of the polymerase activity (Joyce *et al.*, 1994).



**Figure 1.4**

Crystal structure of *E.co* Pol I (PDB ID: 1KLN) in blue. The primer strand (green) can be seen in close proximity to D355 and E357 (red) in the 3'-5' exonuclease domain (Joyce *et al.*, 1994). The template strand can be seen in yellow.

## Sequence motifs

Conserved regions of *polA* sequences were first established by Delarue *et al.* (1990); suggesting polymerase enzymes may share a common tertiary fold, or at least contain similar local tertiary architecture required for similar functions. It was suggested these motifs were likely to represent 'modules' required for the polymerase structure and activity.

Family A DNA polymerases share six motifs that are evolutionary conserved (Figure 1.5). These motif residues contact the substrates and constitute the active site. The most highly conserved are Motifs A, B and C. Motifs 1, 2 and 6 are conserved structurally but can vary in amino acid sequence (Patel *et al.*, 2001).

### *Motif A*

Motif A is one of two conserved motifs present in all DNA and RNA polymerases. Residues in Motif A are located in the Palm subdomain and make contact with the primer strand bases, the sugar-phosphate backbone, and the catalytic metal ions (Derbyshire *et al.*, 1991).

### *Motif B*

The residues of Motif B are located in the Fingers subdomain, and form the O-helix in *T.aq* Pol I. This O-helix is important in that it makes contact with the nascent base-pair during polymerisation and forms hydrogen bonds with the triphosphate of the incoming nucleotide (Steitz *et al.*, 1995).

### *Motif C*

Like Motif A, the amino acid residues of Motif C are located in the Palm subdomain and are highly conserved in all DNA and RNA polymerases. It contains the aspartic residue (Asp<sup>882</sup> in *E.co* Pol I) that coordinates the catalytic metal ions and has been seen to be crucial for polymerase activity (Joyce *et al.*, 1994).

### *Motif 1*

Consisting of a helix and loop, and located at the tip of the Thumb subdomain, Motif 1 interacts with the sugar-phosphate backbone of the DNA template and primer, four to seven base-pairs upstream from the active site (Derbyshire *et al.*, 1991).

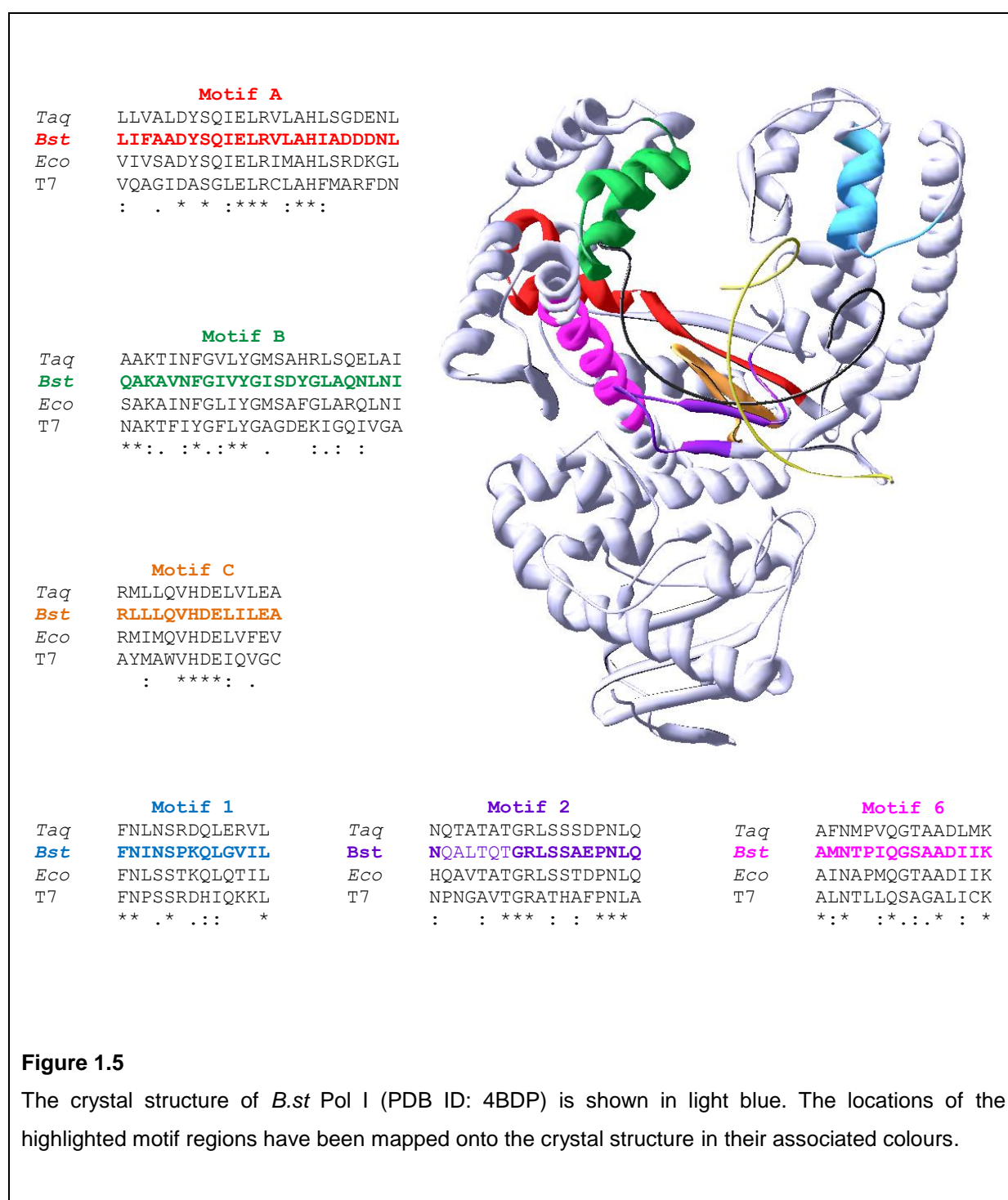
### *Motif 2*

Motif 2 consists of two beta strands located in the Palm subdomain. Amino acid residues in this motif have been observed to interact with the DNA minor groove and the template

sugar-phosphate backbone (Derbyshire *et al.*, 1991).

### Motif 6

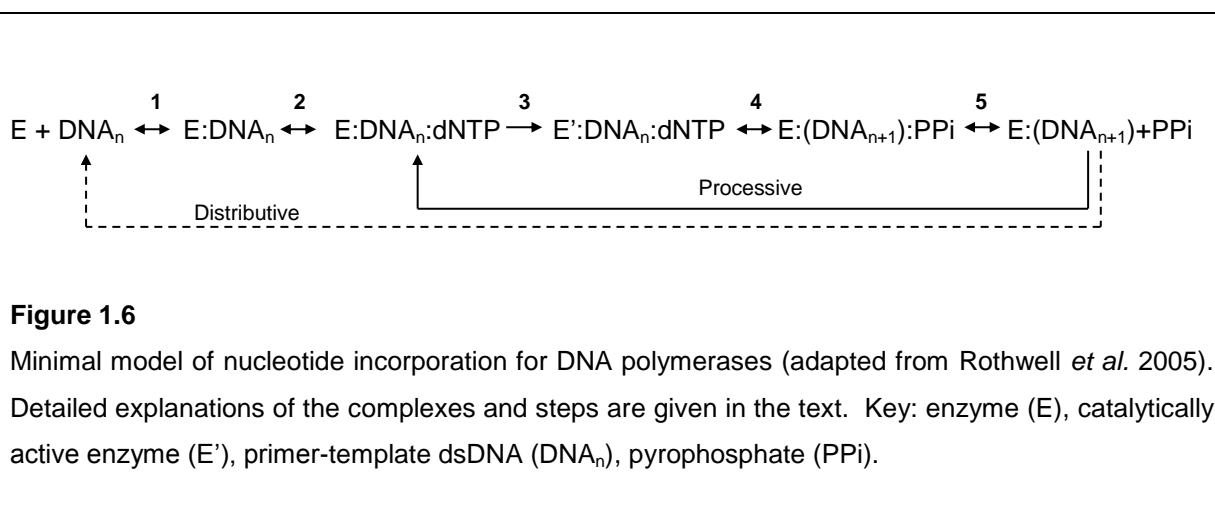
Motif 6 consists of a helix running parallel to the DNA template in the Palm, at the base of the Fingers subdomain. Residues in this Motif participate in binding the DNA in the minor groove (Patel *et al.*, 2001).





### 1.1.2 Nucleotide incorporation

Despite limited sequence identity and differences in overall structure, polymerases have adopted a common mechanism for the essential process of nucleotidyl transfer (Figure 1.6).



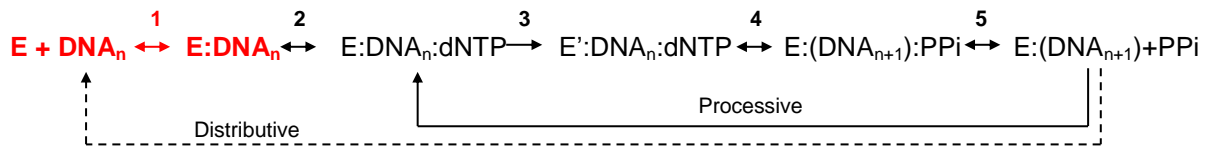
**Figure 1.6**

Minimal model of nucleotide incorporation for DNA polymerases (adapted from Rothwell *et al.* 2005). Detailed explanations of the complexes and steps are given in the text. Key: enzyme (E), catalytically active enzyme (E'), primer-template dsDNA ( $DNA_n$ ), pyrophosphate (PPi).

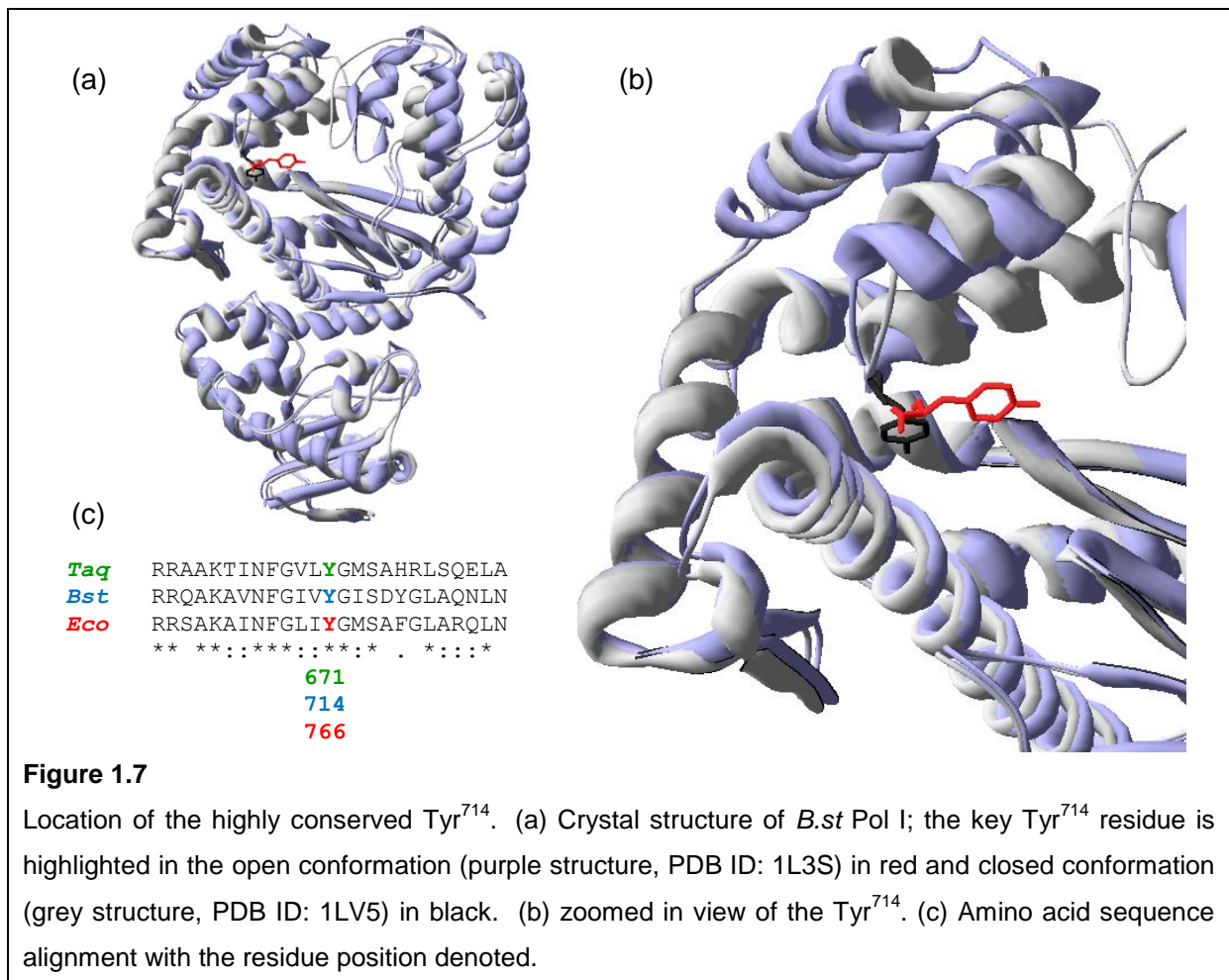
A wide range of crystallographic and kinetic studies on family A DNA polymerases have helped to understand how the enzymes synthesise DNA with exceptionally high accuracy. Polymerase crystal structures in the presence of DNA and dNTP enabled the identification of the exact amino acid residues within the dNTP binding pocket of the catalytic Palm subdomain. Studies on *E.co* Klenow Pol I (Beese *et al.* 1993), T7 Pol I (Doublié *et al.* 1998), *B.st* DNA Pol I (Kiefer *et al.* 1998) and *T.aq* Pol I (Li *et al.* 1998) helped to establish a minimal model (Figure 1.6) of the stages of nucleotide incorporation for DNA polymerases as a whole (reviewed by Rothwell *et al.* 2005) and is described in detail below.



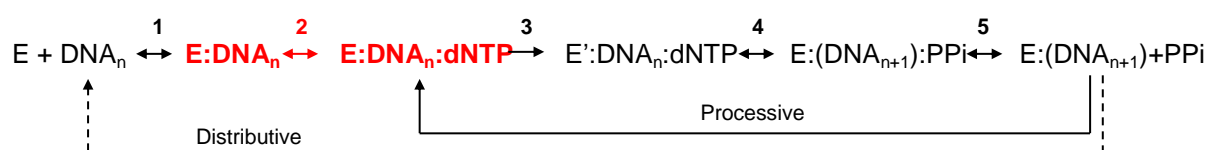
# 1. Primer/template DNA binding and recognition:



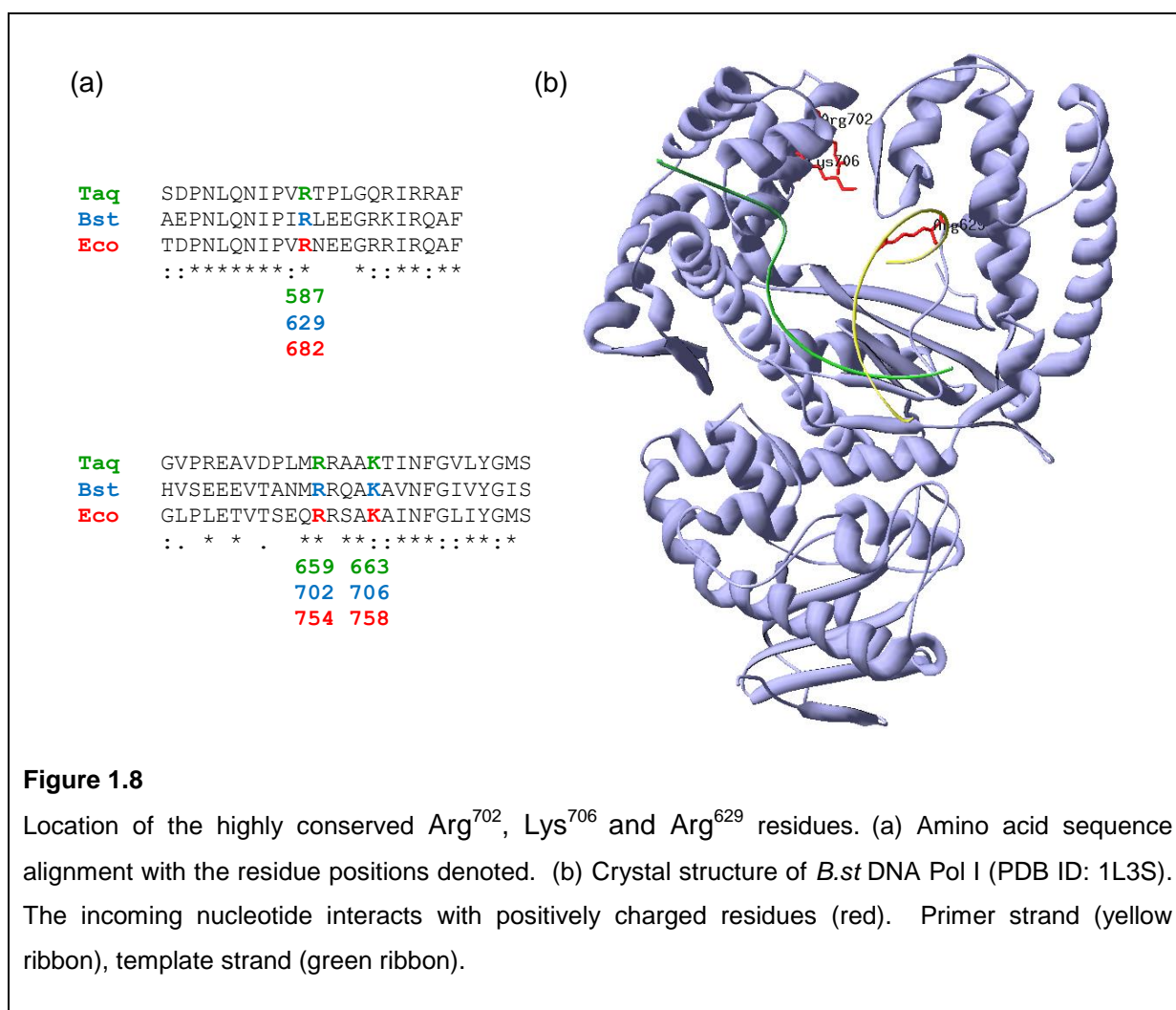
The first stage in the nucleotide incorporation cycle is for the polymerase enzyme (E) to bind the double-stranded primer/template DNA (DNA<sub>n</sub>) to form the enzyme-primer/template complex (E:DNA<sub>n</sub>). DNA binding results in structural changes in the Thumb subdomain (Li *et al.* 1998). This conformational movement of the Thumb surrounds the DNA enabling the active site residues in the Palm subdomain to interact with the template strand at the 3' terminus of the primer. A highly conserved Tyr<sup>714</sup> (*B.st* Pol I) (Figure 1.7) positions the primer/template, ensuring the first base-pair of the duplex DNA can be located in the active site. Interactions within the Palm and Thumb subdomains bend the bound DNA so that the single-stranded region of the template is flipped out of the stacking arrangement at a sharp angle, thereby directing the ssDNA template out of the active site and over the Fingers subdomain (Kiefer *et al.*, 1998).



## 2. Formation of the E:DNA<sub>n</sub>:dNTP complex:



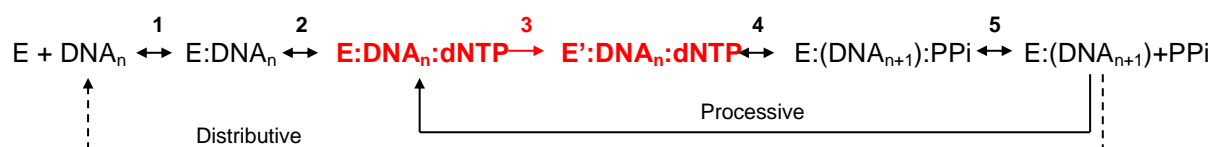
Next, the nucleotide base is incorporated into the enzyme-primer/template complex to form the complex E:DNA<sub>n</sub>:dNTP. The efficiency of incorporation varies between different polymerases and it is this binding step that enables discrimination between correct and incorrect nucleotides. Replicative polymerases show greater efficiency against an incorrect nucleotide incorporation whereas repair enzymes show less (Waksman *et al.*, 2005). Initial recognition of the incoming nucleotide is achieved through interactions with positively charged residues (Arg<sup>702</sup>, Lys<sup>706</sup> and Arg<sup>629</sup> in *B.st* Pol I) (Figure 1.8). When bound, a large conformational change in the Fingers subdomain delivers the nucleotide to the active site in the Palm subdomain (Beese *et al.*, 1993; Li *et al.*, 1998; Rothwell *et al.*, 2005).



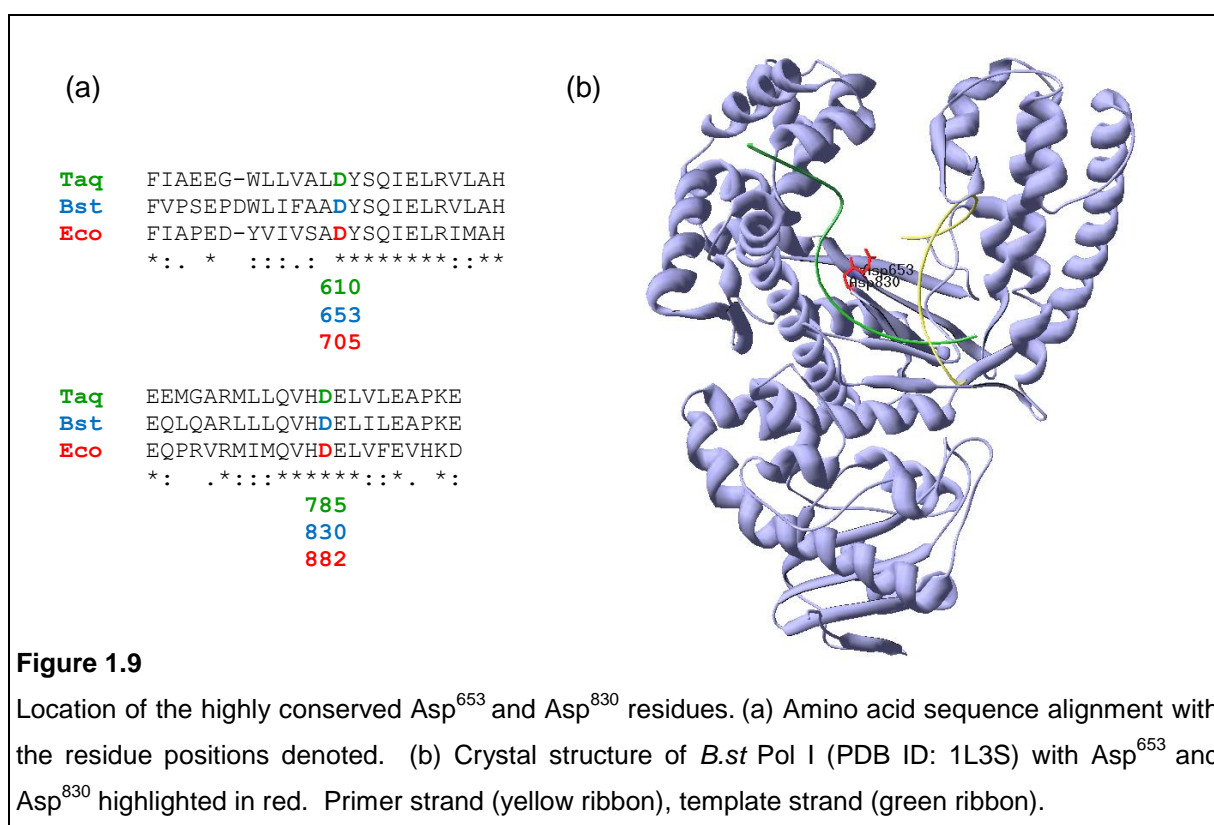
**Figure 1.8**

Location of the highly conserved Arg<sup>702</sup>, Lys<sup>706</sup> and Arg<sup>629</sup> residues. (a) Amino acid sequence alignment with the residue positions denoted. (b) Crystal structure of *B.st* DNA Pol I (PDB ID: 1L3S). The incoming nucleotide interacts with positively charged residues (red). Primer strand (yellow ribbon), template strand (green ribbon).

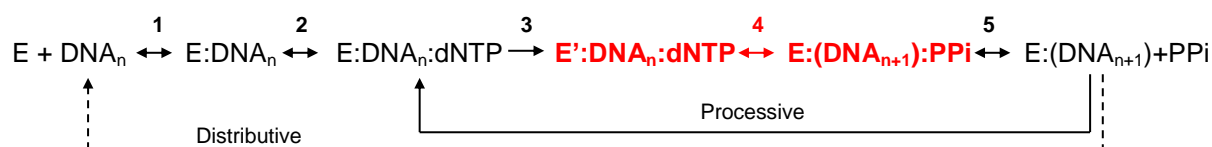
### 3. Transition to a catalytically active E':DNA<sub>n</sub>:dNTP complex:



Only correct nucleotides remain bound long enough for the formation of the ternary E':DNA<sub>n</sub>:dNTP complex that is required for catalysis. It is the conversion to this complex that is thought to be the rate-limiting step of polymerisation and the correct geometric selection ensures accurate incorporation (Goodman *et al.*, 1998). Movement of the O-helix orientates the subdomain from an 'open' to a 'closed' state. In an open state the stacking of a highly conserved Tyr<sup>714</sup> (*B.st* Pol I) with the template base prevents entry of the incoming nucleotide. Orientation to the closed state releases the stacking arrangement, allowing entry of the incoming nucleotide. With correct base pairing, the closure of the Fingers subdomain results in the formation of a pocket that accommodates the nascent base pair. Nucleotide selection within the active site is the major contributor to the fidelity of synthesis (Kunkel *et al.*, 2000). In the closed conformation, the incoming dNTP is bound to the O-helix, stacked onto the 3' base of the primer strand and bound by two metal ions (Mg<sup>2+</sup>) to the catalytic aspartate residues (Asp<sup>653</sup> and Asp<sup>830</sup> in *B.st* Pol I) ready for catalysis (Kiefer *et al.*, 1998).

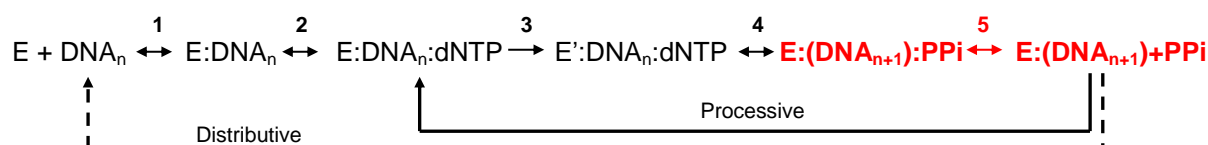


#### 4. Phosphoryl-transfer reaction:



DNA synthesis is mediated by the transfer of a phosphoryl group from the incoming dNTP to the DNA 3'-OH, liberating a pyrophosphate (PPi) and forming a new DNA phosphodiester bond. Two metal ions catalysing the event are coordinated by the triphosphate of an incoming nucleotide and by the highly-conserved acidic residues, usually aspartate, of the Palm subdomain (Derbyshire *et al.*, 1991). The first metal ion activates the 3'-OH of the DNA primer for attack of the incoming nucleotide alpha-phosphate. The second metal ion contacts all three phosphates of the bound nucleotide stabilising the displaced pyrophosphate moiety (Joyce *et al.*, 1998; Steitz, 1999). This conservation of metal-binding sites in highly divergent DNA polymerases demonstrates the importance of the metal ions for assisting nucleotide polymerisation (Rothwell *et al.*, 2005).

#### 5. Product release and translocation of primer/template DNA



All polymerases must translocate down the template strand after the incorporation of the nucleotide (Joyce and Steitz, 1994). A second conformational change in the Fingers subdomain allows the release of the PPi product. The release of the pyrophosphate removes the stabilising interactions with the  $\text{Mg}^{2+}$  ions leading to a reversal of the Fingers conformational change to form the open complex structure (Joyce and Steitz, 1994). In an open state, the polymerase may then either dissociate from the p/t (distributive synthesis) or move along the substrate to form a new 3' terminus for a new round of incorporation (processive synthesis). The Tyr<sup>714</sup> (*B.st* Pol I) residue of the Fingers subdomain is able to probe the environment around the newly-formed base-pair and insert itself on top of it when the p/t DNA has moved away from the active site. Tyr<sup>714</sup> stacking against the newly-formed base-pair stabilises the DNA, permitting another cycle of nucleotide incorporation (Rothwell

*et al.*, 2005). If a non-Watson-Crick base-pair is incorporated, the duplex DNA is destabilised, facilitating the formation of a partially single-stranded 3' terminus capable of moving to the editing exonuclease active site (Beese *et al.*, 1993; Joyce and Steitz, 1994). Alternatively, the polymerase can extend past the mis-incorporated base, 'sealing' it within the elongated strand as a mutation (Rothwell *et al.*, 2005). In either case, it is the rotation of the O-helix of the Fingers subdomain that results in translocation along the template DNA (Joyce and Steitz, 1994).

### 1.1.3 Strand-displacement activity

A feature of some DNA polymerases is the ability to strand-displace downstream DNA. A number of Family A and B polymerases use this activity to unwind duplex DNA without the need for additional protein cofactors such as helicases or single-stranded binding proteins (Hamilton *et al.*, 2001).

Strand-displacement synthesis is an essential process in the removal and replacement of the RNA primers of Okazaki fragments (Kornberg, 1980). Several studies have attempted to localise the intrinsic strand-displacement activity of DNA polymerases. Soengas *et al.* (1992) mutated a highly-conserved aspartic acid residue in the 3'-5' exonuclease domain of the Family B bacteriophage Phi29 Pol, inactivating exonuclease activity and reducing strand-displacement activity. This led to the suggestion that strand-displacement activity resides in the N-terminal domain of the enzyme. However, Derbyshire *et al.* (1988) and Perler *et al.* (1996) showed the complete knock out of the 3'-5' exonuclease activity does not remove the intrinsic strand-displacement activity in either *P.fu* Pol II or *E.co* Pol I. To date, no available DNA polymerase crystal structure contains sufficiently long single-stranded DNA template overhang or downstream double-stranded DNA to provide a detailed explanation or exact localisation of strand-displacement activity. However, crystal structures of Family A polymerase enzymes (T7, *E.co*, *B.st* and *T.aq*) (Doublié *et al.*, 1998; Beese *et al.*, 1998; Kiefer *et al.*, 1997; Li *et al.*, 1998) in the presence of short stretches of primer/template DNA offer potential insights into the location of strand-displacement activity. Singh *et al.* (2007) used these structures in an attempt to further elucidate the residues and motifs responsible for strand-displacement activity.

A key residue located in the Fingers domain, Tyr<sup>714</sup> (*B.st* Pol I), has been mentioned earlier with regard to nucleotide incorporation. This highly-conserved residue occupies the catalytic site binding pocket until the incoming nucleotide is selected. Here, the immediate unpaired template nucleotide assumes a flipped conformation (Li, *et al.*, 1998) and the n+1 nucleotide

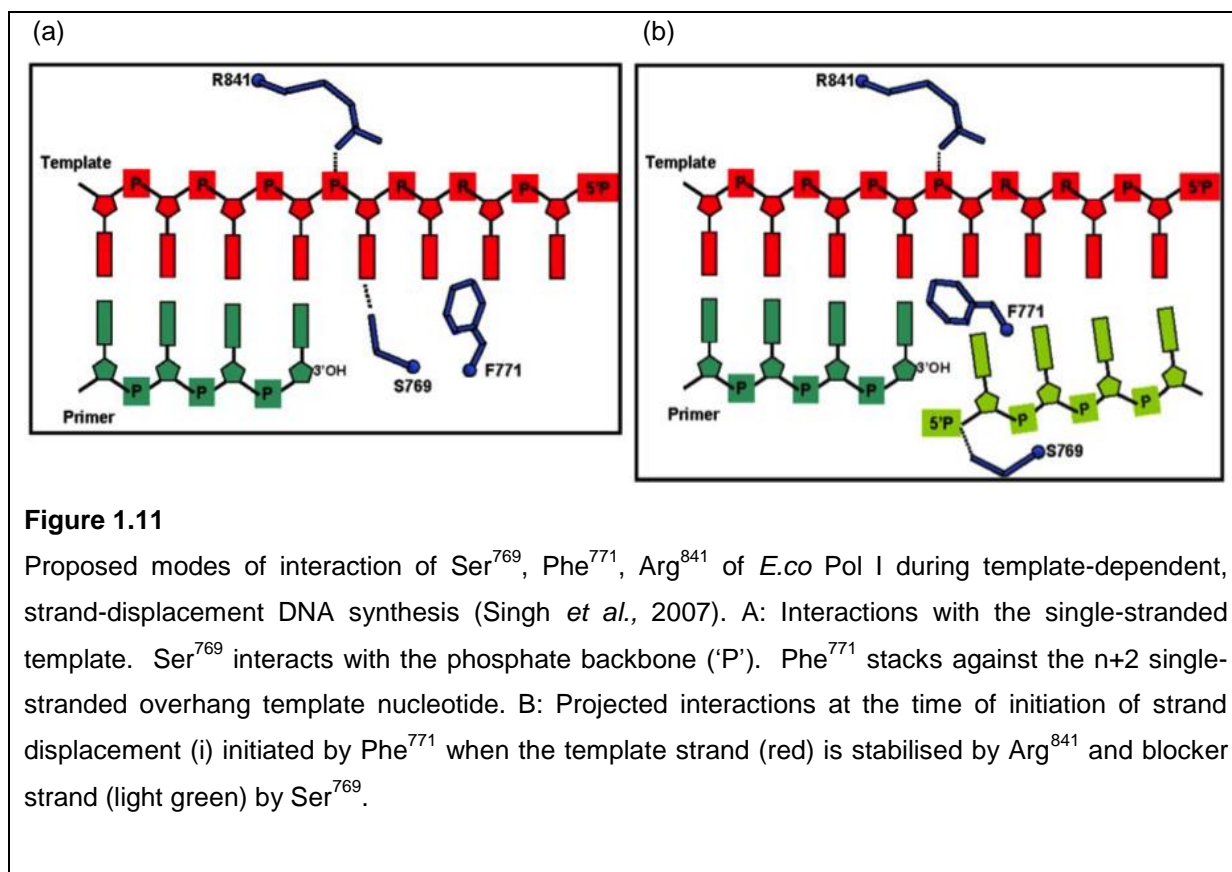


### Figure 1.10

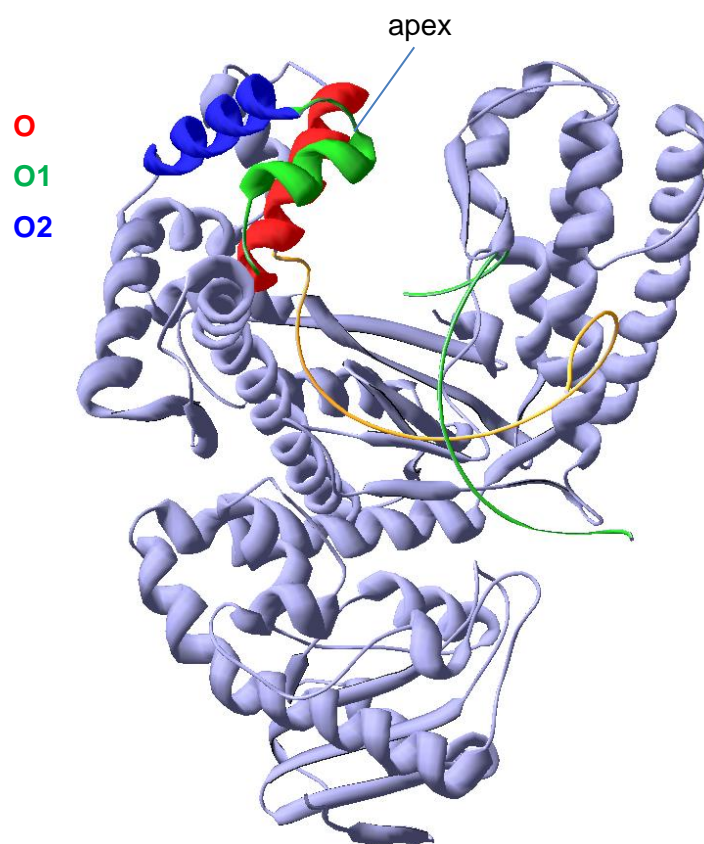
24

Tyr<sup>719</sup> of *B.st* Pol I has been shown to pack against the n+2 nucleotide of the template DNA (Kiefer *et al.*, 1997). Crystal structures of T7 Pol I (Steitz *et al.*, 2004) position the Phe<sup>644</sup> (structurally homologous to Phe<sup>771</sup> *E.co* Pol I, and ) at the DNA fork, implying its role in strand separation.

The crystal structures available directed Singh *et al.* (2007) to target specific residues. Using the strand-displacing *E.co* Pol I enzyme they mutated the homologous Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup> to alanine residues to investigate the strand-displacement effects at each position, as single, double and triple mutants. Using primer extension and primer gap-filling assays to monitor polymerase activity, they noted S769A and F771A, as single or double mutants, retain polymerase activity. However, the double mutant showed reduced polymerase activity, suggesting an inability to displace the non-template strand, a pre-requisite for nucleotide incorporation. DNA synthesis in the presence of a blocking primer enabled the comparison of strand-displacement activity of the mutants (Figure 1.11). F771A and S769A/F771A were shown to be defective in displacement activity, implying the participation of both residues in the generation of the flap-structured DNA, required for strand-displacement. They further showed the R841A mutant retained polymerase activity but removed strand-displacement activity.



Joyce *et al.* (1995) first identified the Fingers subdomain to bind the ssDNA template across from and beyond the site of synthesis and inferred to it a role in template fixation and template specificity. Ser<sup>769</sup> and Phe<sup>771</sup> are located on the O-helix of Pol I as defined by Ollis *et al.* (1985). The O-helix bundle is made up of O, O1 and O2 helices (Figure 1.12). This 3-helix bundle has further been shown to share structural homology with the DNA binding motif, Mrf-2 (Yuan *et al.*, 1998). Singh *et al.* (2007) have suggested the motif is a functional unit that recognises or induces altered DNA structure and requires an aromatic residue at its apex for strand separation.



**Figure 1.12**

Crystal structure of *B.st* Pol I (PDB ID: 3BDP) with the O-helix bundle highlighted on the Fingers subdomain. The apex defines the position of the aromatic residue suggested to be a requirement for strand separation. O-helix (red), O1-helix (green), O2-helix (blue).

Sequence alignments show the three residues to be highly conserved among known strand-displacing polymerases (Figure 1.10c). Singh *et al.* (2007) have therefore suggested the presence of Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup> in Pol I enzymes to be a prerequisite for strand-displacement activity.



This strand displacement activity is not confined to the bacterial world. In Eukarya, Pol  $\delta$  has been shown to displace DNA strands (Maga *et al.*, 2001) and in the Archaea, *Pyrococcus abyssi* (*P.ab* PolD) (Henneke *et al.*, 2005) and *Sulfolobus solfataricus* (*S.so* PolB1) (Huang *et al.*, 2013) have the ability to strand-displace. However, the strand displacement activity and processivity of replicative polymerases is enhanced by accessory proteins that topologically surround their DNA substrates and are often essential to enable activity (Joyce and Steitz, 1995).

## **1.2 DNA amplification methods**

Nucleic acid amplification is one of the most valuable tools in biotechnology. Nucleic acid testing (NAT) techniques target DNA sequences directly. Several molecular methods have been developed to overcome the shortcomings of the classical diagnosis methods. In this way, amplification of nucleic acids is widely used in forensics, medicine, and agriculture. The classical method of microbial detection and identification remains the culture and subsequent phenotypic differentiation of the causative pathogen. Direct observation of microbes by microscopy is employed as a simple diagnosis method enabling cost-efficient use in resource-limited laboratories in developing countries. However, slow growth rates often limit culture-based diagnosis and delays effective treatment. The establishment of more rapid, sensitive and accurate diagnosis methods is required to replace the time-consuming and limited culture methods (Mori *et al.*, 2009; Craw *et al.*, 2012).

### **1.2.1 Nucleic acid detection**

The polymerase chain reaction (PCR) is now a well-established, rapid and highly-sensitive method of *in vitro* DNA amplification (Mullis *et al.*, 1987). The reaction comprises three stages: (i) double-stranded DNA is denatured by a high temperature step, (ii) the reaction is cooled to a temperature favourable to specific primer annealing, and (iii) the temperature is raised to that of the optimal temperature for a thermostable polymerase to extend from the bound primer. Amplification occurs through the successive cycling nature of this reaction. PCR is the current gold-standard nucleic acid amplification technique, and has been since the isolation of the thermostable *T.aq* Pol I (Saiki *et al.*, 1988). Real-time PCR has further enabled the quantification of target nucleic acids. Fluorescent chemistries, in combination with highly sensitive optical instrumentation, enables real-time PCR to offer accurate, rapid, quantitative measurement, with high sensitivity and specificity.

However, with the development of new amplification techniques, PCR can now be seen to have several disadvantages. The precision thermal cycling required restricts its application to large, power-demanding instrumentation, which in turn requires it to be performed within a laboratory, often with trained technicians to perform the reactions. Consequently, a variety of non-PCR 'isothermal' techniques have been developed, removing the requirement for a thermo-cycling machine altogether. Each of these amplification methods has its own innovation to promote the next round of DNA synthesis. The simplicity and isothermal nature of these methods offer great potential for the development of low-cost, low-power, hand-held devices used to detect pathogens, for example, at the point-of-care or point-of-sampling.

### **1.2.2 Isothermal nucleic acid amplification**

Isothermal DNA amplification techniques vary in their levels of sensitivity, reaction complexity, amplification speed and specific reagent requirements, and have been reviewed extensively (Gill *et al.*, 2008; Asiello *et al.*, 2011; Niemz *et al.*, 2011; and Craw *et al.*, 2012).

Popular isothermal methods include: rolling circle amplification (RCA) (Fire *et al.*, 1995), strand-displacement amplification (SDA) (Walker *et al.*, 1998), nucleic acid sequence-based amplification (NASBA) (Compton, 1991), Helicase-dependant amplification (HDA) (Vincent *et al.*, 2004), and Loop-mediated isothermal DNA amplification (LAMP) (Notomi *et al.*, 2000). A comparison of the common isothermal methods are presented in Table 1.2.

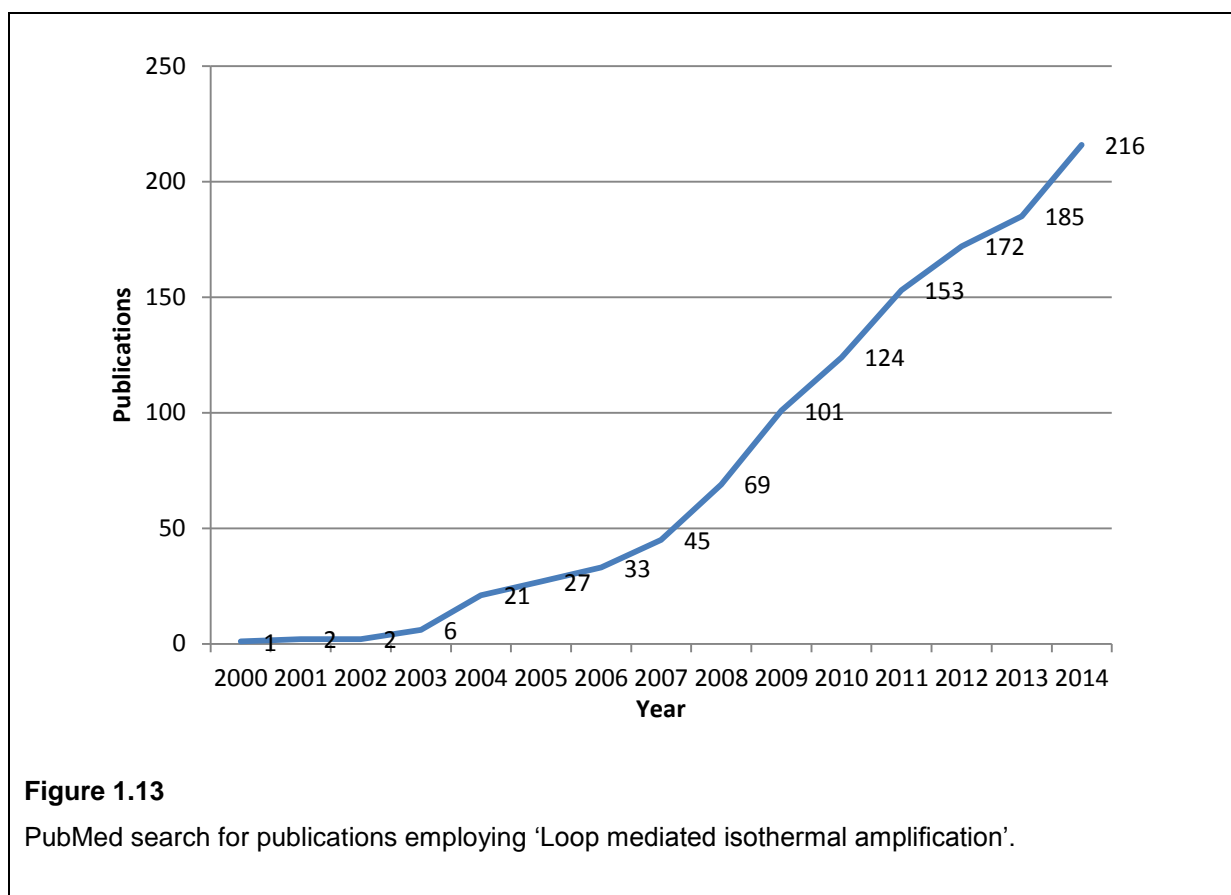
HDA uses a helicase to physically separate the strands of double-stranded DNA, permitting primer binding and extension by a DNA polymerase at 37°C or 65°C (Vincent *et al.*, 2004). NASBA is a method for the isothermal amplification of RNA based on transcription (Compton, 1991). The inherent RNA selectivity of NASBA enables the method to be used specifically to detect viable cells; mRNA is less stable than DNA and therefore degrades rapidly in dead cells (Bentsink *et al.*, 2002). RCA is used to amplify circular nucleic acids using the strand-displacement activity of Phi29 DNA polymerase (Fire *et al.*, 1995).

	PCR	LAMP	HDA	SDA	RCA	NASBA
Preferred template	dsDNA	ssDNA	dsDNA	ssDNA	Circular ssDNA	RNA
Amplicon type	dsDNA	Concatenated dsDNA	dsDNA	dsDNA	Concatenated ssDNA	RNA
Reaction temperature (initial temp) °C	95, 55-60, 72 (95)	60-65 -	37-65 -	37 (95)	37 (95)	37-42 (95)
Primers required	2	4 + 2 loop primers	2	2 (DNA/RNA chimeric) + 2 bumper primers	1	2
Primer design	Simple	Complex	Simple	Complex	Simple	Simple
Enzymes required	1	1	2	2	1	2-3
Denaturing agent	Heat	Betaine / Strand-displacing-polymerase	Helicase	Restriction enzymes	Strand-displacing polymerase	RNase H
Performance	>120 mins	<60 mins	<100 mins	<120 mins	>90 mins	<60 mins
Sensitivity	<10 copies	>10 copies	1 copy	>10 copies	>10 copies	1 copy
Specificity	High	Very high	High	High	High	High
Detection method	Gel electrophoresis Real-time fluorescence	Gel electrophoresis Turbidity Visible colour change Real-time fluorescence	Gel electrophoresis Real-time fluorescence	Gel electrophoresis Real-time fluorescence	Gel electrophoresis Real-time fluorescence	Gel electrophoresis Real-time fluorescence
Multiplex	Yes	Yes (limited)	Yes	-	Yes	Yes
Tolerance to inhibitors	Poor	Good	Good	Poor	Poor	Poor
Key publication	Mullis (1987)	Notomi (2000)	Vincent (2004)	Walker (1992)	Fire (1995)	Compton (1999)

**Table 1.2** Comparison of nucleic acid amplification methods.

### 1.2.3 Loop-mediated isothermal amplification

Since loop-mediated isothermal amplification (LAMP) was originally reported by Notomi *et al.* (2000), >1160 papers employing LAMP have been published (Figure 1.13). This is in contrast to the limited publications of alternative isothermal DNA amplification methods. As such, LAMP appears to be the increasingly preferred isothermal method of choice due to its rapid, accurate, cost-effective and reliable methodology.

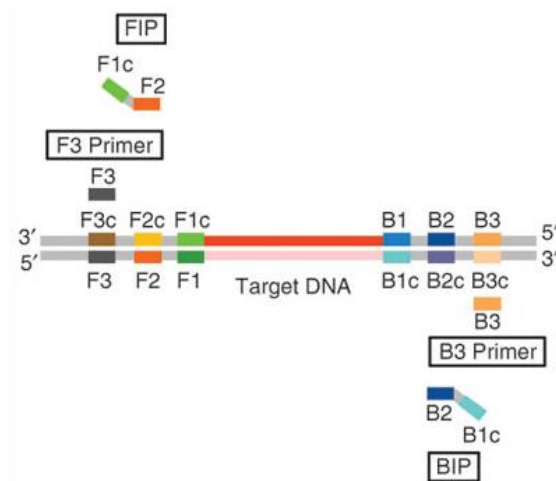


A typical LAMP reaction requires a set of four to six specific primers and a DNA polymerase with strand-displacement activity. The Large Fragment Pol I enzyme from *Geobacillus stearothermophilus* (*B.st* LF Pol I) is routinely used, as its inherently strong strand-displacement activity acts as a helicase to unwind the double-stranded DNA. LAMP reactions are performed at 65°C, the optimum reaction temperature for *B.st* LF DNA Pol I. A further, and significant, advantage of using a high isothermal temperature, is that the target specific primers must also anneal at this temperature, restricting non-specific amplification of false products. The LAMP final products are a mixture of stem-loop DNAs with various stem lengths and cauliflower-like structures, with multiple loops formed by annealing between alternatively inverted repeats of the target sequence in the same strand (Notomi *et al.* 2000).

LAMP has proved to be a highly-efficient amplification technique and can generate microgram quantities of template DNA in a very short time (Notomi *et al.*, 2000). An overview of the highly complex reaction pathway is summarised below and is adapted from Mori *et al.* (2008), Notomi *et al.* (2000), and the Eiken Chemical Co. Ltd. website (<http://www.loopamp.eiken.co.jp/e/lamp/>).

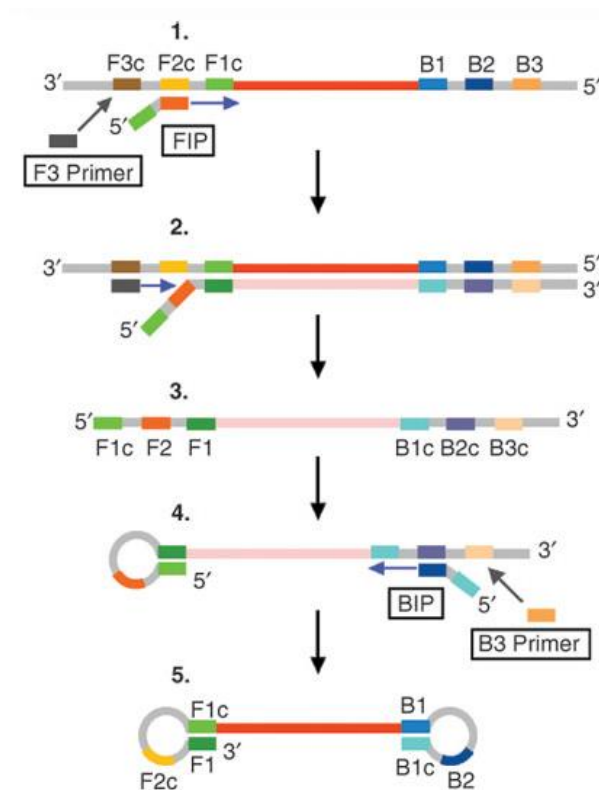
## LAMP method:

### (A) Primer design of the LAMP reaction



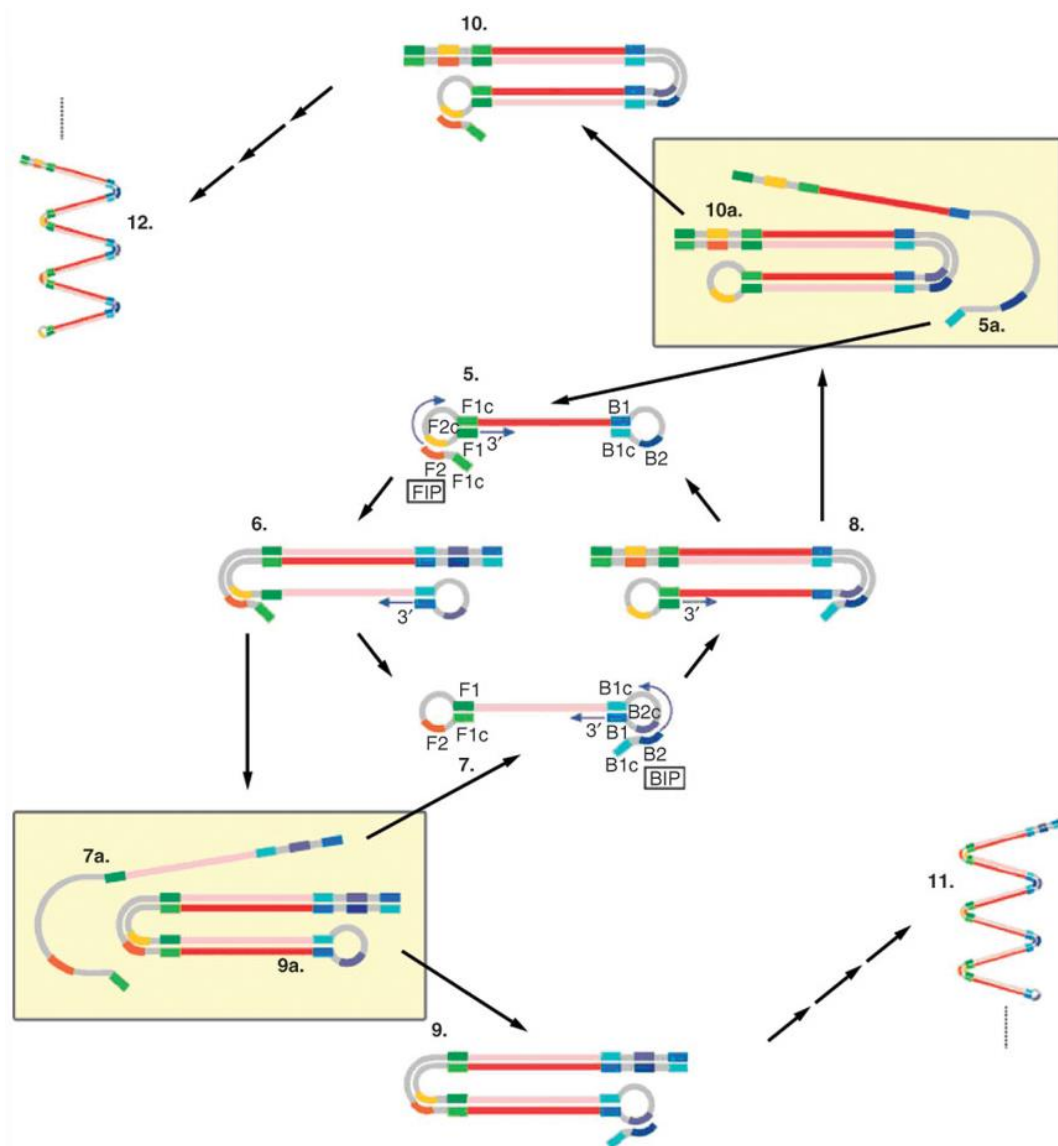
- Six distinct regions are designated on the target DNA, labelled F3, F2, F1, B1c, B2c and B3c from the 5' end, with 'c' representing a complementary sequence. For example, the F1c sequence is complementary to the F1 sequence.
- Two inner primers (FIP and BIP) and outer primers (F3 and B3) are used in the LAMP method. FIP (BIP) is a hybrid primer consisting of the F1c (B1c) sequence and the F2 (B2) sequence.

(B) Starting structure/Dumbbell formation step:



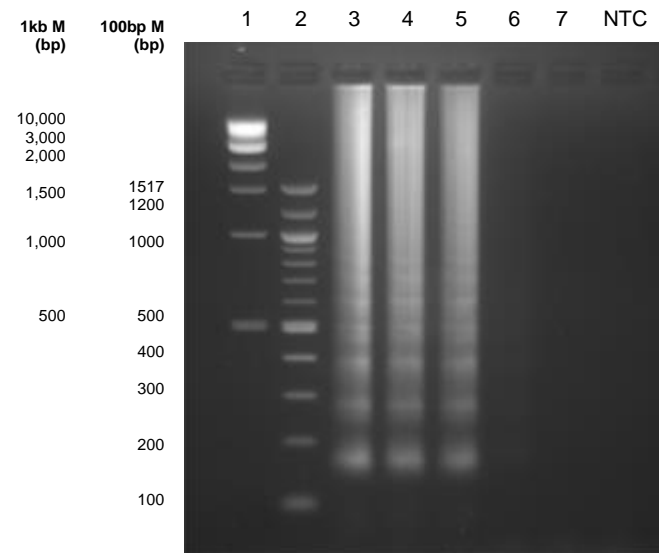
- DNA synthesis is initiated from the FIP primer.
- The F2 region anneals to the F2c region on the target DNA and initiates the elongation (1). DNA amplification proceeds with BIP in a similar manner.
- The F3 primer anneals to the F3c region on the target DNA, and strand-displacement DNA synthesis takes place (2). The DNA strand elongated from FIP is therefore replaced and released (3).
- The released single strand forms a loop structure (4). DNA synthesis proceeds with the single-strand DNA as the template, and BIP and B3 primers, in the same manner as described earlier, to generate (5), which possesses the loop structure at both ends (dumbbell-like structure).

(c) *Cycling amplification step:*



- Using self-structure 5 as the template, self-primed DNA synthesis is initiated from the 3' end F1 region, and the elongation starts from FIP annealing to the single strand of the F2c region in the loop structure.
- Elongation and strand-displacement events generate a variety of amplicon repeat structures.
- Structure 7 is generated, which is complementary to structure 5, and structure 5 is produced from structure 8 in a reaction similar to that which led from structures 5–7. Structures 9 and 10 are produced from structures 6 and 8, respectively, and more elongated structures (11, 12) are also produced.

- All reaction products contain the starting target amplicon, in multiple stem-loop repeats, resulting in a characteristic ladder-like appearance when visualised by gel electrophoresis (Figure 5.14).



**Figure 1.14**

A 2% (w/v) agarose gel showing the typical banding patterns generated from a LAMP reaction. Lane1: 500ng NEB 1kb ladder, lane2: 500ng NEB 100bp ladder, lane 3-7: LAMP reaction product amplified from  $1 \times 10^{5.4,3,2,1}$  copies *Aeropyrum pernix* (*A.pe*) gDNA, lane 8: NTC.

LAMP does not require initial template denaturation (Nagamine *et al.*, 2001), but a brief incubation at 95°C, to separate the starting template, has been found to improve amplification times (Suzuki *et al.*, 2010). If an initial heat step is used, the *B.st* Pol I must be added afterwards, to avoid enzyme denaturation. This additional manipulation poses a contamination risk and would be best avoided. The inclusion of two additional primers termed 'loop primers' has been shown to accelerate amplification further, by providing additional sites for strand-displacement and primer extension. The use of loop primers has also been reported to increase target sensitivity (Nagamine *et al.* 2002).

LAMP can be modified for the detection of RNA, i.e. RT-LAMP, simply by the addition of a specific reverse transcriptase enzyme to the reaction. LAMP has also been demonstrated to be more tolerant to common PCR inhibitors (Kaneko *et al.*, 2007), potentially enabling LAMP



to be used with simplified nucleic-acid extraction methods, further increasing the method's ease of use and application to point-of-care.

#### **1.2.4 Applications for Biotechnology**

DNA polymerases have become an essential *in vitro* diagnostic and analytical tool for the molecular biologist. Polymerase activities have been harnessed for applications such as amplification, labelling and detection of DNA sequences. Enzyme requirements including nuclease activity, thermostability, processivity, strand-displacement activity, fidelity and the ability to incorporate modified nucleotides are desired for some roles but not for others. The polymerases from the Family A (Pol I) and Family B (Pol II) have the most value in the biotechnology sector. These enzymes have been investigated extensively because of their simplicity in subunit composition, their ease of cloning, over-expression and purification, and their detailed biochemical characterisation (Hamilton *et al.*, 2001).

The intrinsic proofreading activity, high processivity and strand-displacement activity of bacteriophage Phi29 DNA Pol has enabled its use as a commercial tool for isothermal, rolling-circle DNA amplification (Fire *et al.*, 1995; Liu *et al.*, 1996), and whole genome amplification (Dean *et al.*, 2001). Its use as a commercial enzyme is, however, restricted to applications at or below 30°C due to its low thermostability.

Several moderately thermostable strand-displacing DNA polymerases have been reported, yet *B.st* Pol I remains the enzyme of choice for the majority of isothermal DNA amplification reactions. Few highly thermostable DNA polymerases have been reported to show strand-displacement activity, with none having been demonstrated to be suitable for isothermal DNA amplification reactions, including LAMP.

### 1.3 Project objectives

With the growing importance of isothermal nucleic acid amplification as a means for on-site, sensitive and specific detection, a need has arisen for the identification of alternative DNA polymerase enzymes to enhance existing, and potentially enable as yet unforeseen, applications.

To date, the majority of isothermal nucleic amplification reaction publications, requiring a moderately thermostable strand-displacing DNA polymerase, report the use of the commercial *B.st* LF Pol I (NEB, UK). Several applications can be envisaged where *B.st* LF Pol I is currently restricting the isothermal method directly. Polymerases with enhanced strand-displacement activity or increased nucleotide incorporation activities may offer faster reactions, further reducing time to amplification. It would be advantageous to identify a hyperthermostable strand-displacing DNA polymerase enabling a novel closed-tube, single-step high temperature step before isothermal amplification. This may enable direct sampling, removing the requirement for sample extraction and purification, thereby significantly reducing detection times. A highly thermostable enzyme will allow the closed-tube heat denaturation of the target DNA, permitting high temperature strand separation prior to LAMP; this is envisaged to increase target specificity, reduce the risk of contamination, and further reduce the time to detection.

The aim of this thesis was to identify new Pol I enzymes suitable for LAMP. This was to be achieved through the screening of novel thermophilic organisms, with subsequent cloning, overexpression and purification of their encoded Pol I. The activity of the enzymes were altered and further enhanced by the identification of key motifs and amino acid residues attributed to specific characteristics. This work was accomplished through the site-directed mutagenesis of specific residues and by the formation of chimeras between novel enzymes. Structural modelling and a comparison to previously reported crystal structures of Pol I enzymes was carried out to reveal further motifs and residues for mutation and to increase our understanding of the strand-displacement activity of these specialised enzymes.

## 2. MATERIALS AND METHODS

### 2.1 Materials

Chemicals were purchased from Sigma (Gillingham, UK) unless otherwise stated.

#### 2.1.1 Media

Luria Broth (LB): Tryptone (10g/litre) (LabM, UK), yeast extract (5g/litre) (LabM, UK), NaCl (5g/litre). For LB agar Petri dishes 1.5% (w/v) agar (LabM, UK) was added prior to autoclaving.

Terrific Broth (TB): Yeast extract (24g/litre), tryptone (12g/litre), 0.8% (v/v) glycerol. Medium was made up to 900ml and autoclaved. Once sterilised, 100ml sterile 0.17M  $\text{KH}_2\text{PO}_4$ , 0.72M  $\text{K}_2\text{HPO}_4$  was added prior to use.

SOC: Tryptone (20g/litre), yeast extract (5g/L), 100mM NaCl, 2.5mM KCl. Once autoclaved, 20mM autoclaved  $\text{MgSO}_4$  and 20mM filter-sterilised glucose were added upon use.

YENB: Yeast extract (7.5g/litre), nutrient broth (8g/litre) (LabM, UK), made up to final volume with sterile Milli-Q water.

Kanamycin: 50mg/ml dissolved in water and filter-sterilised using a 0.25 $\mu\text{m}$  filter (Millipore, UK).

Chloramphenicol: 34mg/ml dissolved in ethanol.

IPTG: 1M solution made up in water and passed through a 0.25 $\mu\text{m}$  filter.

X-Gal: 40mg/ml dissolved in dimethylformamide (DMSO).

### 2.1.2 Buffers

gDNA extraction buffer: 10mM Tris-HCl (pH8.0 @ 25°C), 1mM EDTA, 7.5% (w/v), Chelex-100, 1% (w/v) SDS.

Cell lysis buffer: 500mM Tris-HCl (pH8.1 @ 25°C), 50mM NaCl, 0.1mM EDTA.

Sequencing buffer: 400mM Tris base, 10mM MgCl<sub>2</sub>.

Protein loading buffer: 50mM Tris-HCl (pH6.8 @ 25°C) 100mM DTT, 2% (w/v) SDS, 0.1% (w/v) Bromophenol blue (Fisher Scientific, UK), 10% (v/v) glycerol.

SDS PAGE running buffer: 1.5M Tris base, 14.4g/litre glycine, 1g/litre SDS.

SDS PAGE stain: 0.25% (w/v) Coomassie brilliant blue (R250), 50% (v/v) Milli-Q water, 40% (v/v) Methanol (SureChem Products Ltd., UK), 10% (v/v) glacial acetic acid (SureChem Products Ltd., UK).

SDS-PAGE de-stain: 50% (v/v) Milli-Q water, 40% (v/v) Methanol, 10% (v/v) glacial acetic acid.

Resuspension Buffer: 50mM Tris-HCl (pH8.0 @ 25°C), 50mM NaCl, 0.1mM EDTA.

Purification Buffer A: 50mM Tris-HCl (pH8.0 @ 25°C), 500mM NaCl, 5mM Imidazole.

Purification Buffer B: 50mM Tris-HCl (pH8.0 @ 25°C), 500mM NaCl, 1M Imidazole.

Purification Buffer C: 20mM Tris-HCl (pH8.0 @ 25°C) 0.1mM EDTA, 50mM NaCl.

Purification Buffer D: 20mM Tris-HCl (pH8.0 @ 25°C) 0.1mM EDTA, 1M NaCl.

Polymerase Storage Buffer: 50mM Tris-HCl (pH8.3 @ 25°C), 100mM KCl, 0.1mM EDTA, 1mM DTT, 0.5% (v/v) Tween-20, 0.5% (v/v) IGEPAL<sup>®</sup> CA-630, 50% (v/v) glycerol.

10X THERMOPol Buffer (NEB, UK): 200mM Tris-HCl (pH8.8 @ 25°C), 100mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 100mM KCl, 20mM MgSO<sub>4</sub>, 1% (v/v) Triton X-100.

10X iBuffer: 500mM Tris-HCl (pH8.1 @ 25°C), 300mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 300mM KCl, 50mM MgSO<sub>4</sub>, 1% (v/v) Triton X-100.

10X PCR Buffer: 750mM Tris-HCl (pH8.2 @ 25°C), 300mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub>, 300mM KCl, 30mM MgSO<sub>4</sub>, 1% (v/v) Triton X-100.

## **2.2 Methods: Identification**

Oligonucleotide primers were synthesised from Invitrogen (UK) unless otherwise stated.

### **2.2.1 Strain isolation**

Soil samples donated for this study were collected from a number of global thermal features. For each sample, 1g soil was resuspended in 3ml sterile water. A sterile loop was used to transfer 10µl of the resuspension to an LB agar plate for incubation overnight at 55°C. A single bacterial colony was re-streaked and incubated as before to confirm the isolation of a single strain.

### **2.2.2 Genomic DNA extraction**

Genomic DNA (gDNA) was released from the bacterial cells using a modified cell lysis method described by Gotz *et al.* (2002) and Ausubel *et al.* (1994). Liquid culture strains, provided by the Deutsche Sammlung von Mikroorganismen und Zellkulturen GmbH (DSMZ) culture collection (Germany), were microfuged at 13,000xg for 1min to harvest the cells and the cell pellet resuspended in 500µl gDNA extraction buffer. For plated strains, a sterile 10µl loop was used to transfer cells to 500µl resuspension buffer. 20µl Proteinase K (10mg/ml) was added and the tubes inverted to mix. The cell resuspensions were incubated at 50°C for 1h with occasional inverting to mix. Samples were centrifuged at 5,000xg for 2min to pellet cell debris and remove the Chelex-100. The supernatant was transferred to a fresh tube and mixed with 100µl 5M NaCl and 80µl 10% (w/v) CTAB (dissolved in 0.7M NaCl) and incubated for 30min at 60°C. The sample was centrifuged as before and the genomic DNA in the supernatant extracted with phenol/chloroform. An equal volume of buffer-saturated phenol (Fisher Scientific, UK) was mixed with the DNA sample and mixed by vortexing for 1min, then microfuged at 13,000xg for 5min. The upper aqueous layer was removed to a fresh tube and an equal volume of buffer-saturated phenol and chloroform (1:1) added. The

samples were mixed and centrifuged as before. The upper aqueous layer was removed to a fresh tube containing an equal volume of chloroform. The centrifuge step was repeated, and the chloroform step repeated again. The upper aqueous phase was then transferred to a fresh tube and isopropanol precipitated. Isopropanol (0.8 volumes) was added to the DNA sample, mixed, and immediately centrifuged at 13,000xg for 20min. The pelleted DNA was washed twice with 100µl 70% (v/v) ethanol in water and the pellet air dried for 10min. The pellet was resuspended in 50µl Milli-Q water and the purity and yield of the extracted DNA was estimated by running an aliquot on a 1% (w/v) agarose gel stained with ethidium bromide. If insufficient gDNA had been recovered, the gDNA sample was further amplified using a GenomiPhi<sup>TM</sup>V2 DNA amplification kit (GE Healthcare, UK) following the manufacturer's instructions.

### **2.2.3 16S rRNA gene amplification**

Genomic DNA was used as template in a PCR to amplify the 16S rRNA gene to identify strains of unknown species. The 50µl PCR contained 1X MAXA<sup>Taq</sup> PCR Mastermix (GeneSys Biotech Ltd., UK), 25pmol 16S\_Universal\_27F primer and 16S\_Universal\_1429R primer (Appendix i.i) and 10ng gDNA, and was cycled on a Veriti<sup>®</sup> Thermal Cycler (Applied Biosystems, USA) at 95°C for 3min, followed by 45 cycles (95°C – 10s, 55°C – 10s, 72°C – 2min). 5µl PCR product was visualised on a 1% (w/v) agarose gel stained with ethidium bromide (0.5µg/ml) to confirm amplification of the 16S rDNA fragment.

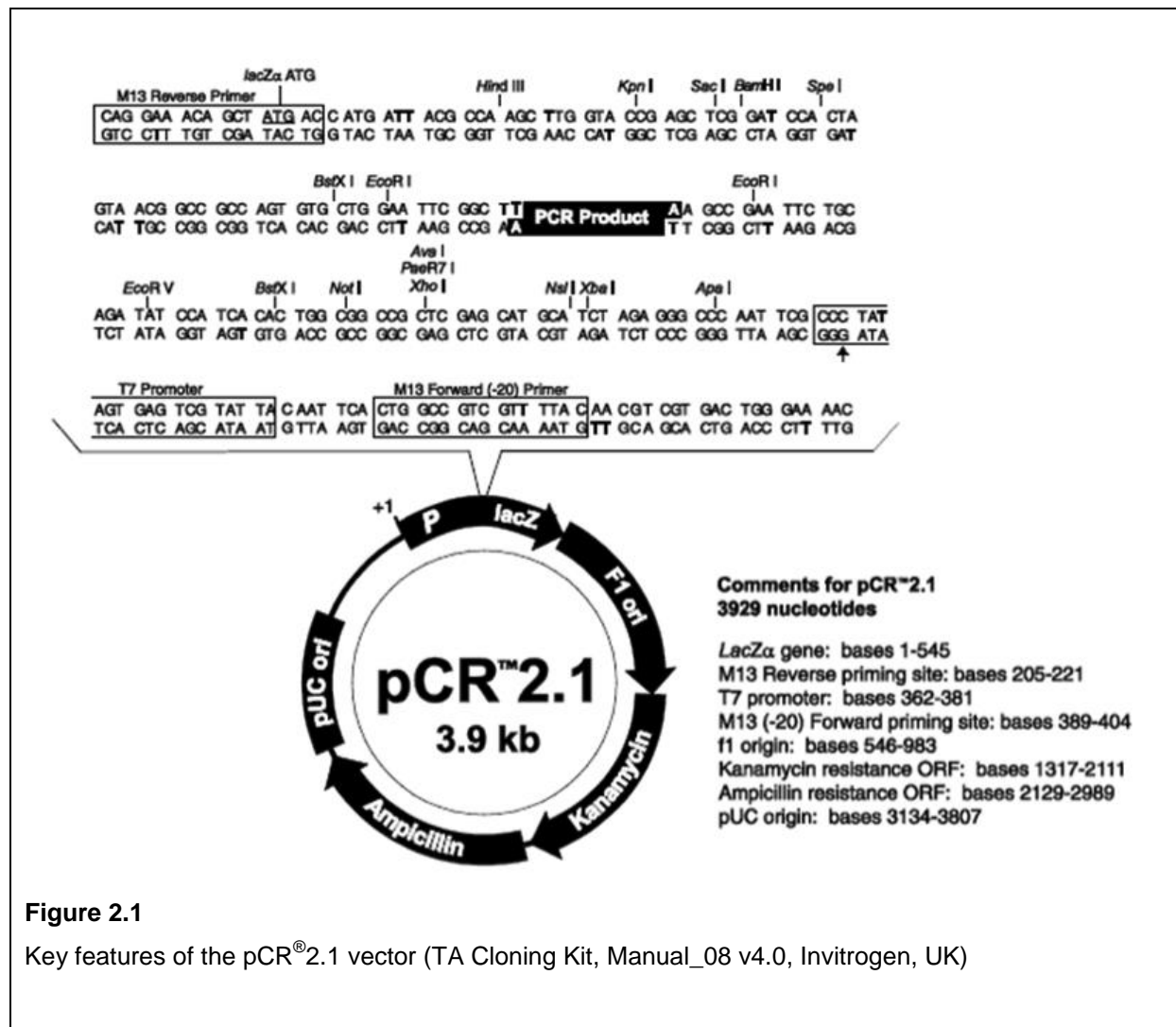
### **2.2.4 *polA* sequence identification**

*polA* fragments were amplified by PCR using degenerate PCR primers (Bergquist *et al.* 2004). The 50µl PCR contained 1X *T.aq* PCR Mastermix (GeneSys Ltd., UK), 25pmol upper primer (PolATF or PolGCF1 or PolGCF2) and lower primer (PolATR or PolGCR), 10ng gDNA, and cycled on an ABI Veriti<sup>®</sup> Thermal Cycler at 95°C – 3min then 16 cycles (95°C – 10s, 60°C – 10s(-1°C/cycle), 72°C – 30s), and then 35 cycles (95°C – 10s, 50°C – 10s, 72°C – 30s). 5µl PCR product was visualised on a 1% (w/v) agarose gel stained with ethidium bromide (0.5µg/ml) to confirm amplification.

### **2.2.5 TA Cloning**

*T.aq* DNA polymerase amplified PCR fragments contain a single non-template-directed deoxyadenosine (dA) residue at the 3' end of duplex PCR products (Clarke, 1988). This overhang enables the direct cloning of the PCR fragment into a vector containing dT

overhangs without the need for restriction endonuclease digestion (Holton, 1990). 1µl PCR product was ligated into 50ng pCR<sup>®</sup>2.1 vector (Original TA Cloning Kit, Invitrogen, UK) (Figure 2.1), as per manufacturer's instructions. Ligations were incubated at 16°C overnight. Reactions were then heated to 70°C for 20min to denature the T4 DNA Ligase present in the sample. The reaction could then be ethanol precipitated for downstream applications.



**Figure 2.1**

Key features of the pCR<sup>®</sup>2.1 vector (TA Cloning Kit, Manual\_08 v4.0, Invitrogen, UK)

## 2.2.6 Ethanol precipitation

DNA was precipitated using 0.1 volumes of 2.5M sodium acetate (pH5.2 @ 25°C) and 2 volumes of 100% (v/v) ethanol. The mix was incubated for 30min at -20°C to precipitate the DNA. The sample was microfuged for 10min at 13,000xg and the supernatant discarded. The DNA pellet was then washed with 100µl 70% (v/v) ethanol in water and centrifuged as before. The supernatant was discarded and the pellet air dried for 10min. The DNA pellet was resuspended in an appropriate volume of Milli-Q water.

### 2.2.7 Transformation

Plasmid DNA was electroporated into *E.coli* competent cells. pCR<sup>®</sup>2.1 constructs were transformed into *E.coli* TOP10F' (Life Technologies, UK) suitable for cloning (Table 2.1). pET vectors (Novagen, UK) were transformed into *E.coli* KRX (Promega, UK) (Table 2.1) that had been modified to introduce an additional chloramphenicol (*Cr*) resistant plasmid pRARE2 (previously isolated from *E.coli* Rosetta<sup>™</sup>2 (Novagen, UK). The pRARE2 plasmid encodes 7 minor tRNAs (AUA, AGG, AGA, CUA, CCC, GGA, CGG) in *E.coli*. The addition of the pRARE2 was expected to aid expression by overcoming limitations in codon usage of the *E.coli* KRX strain. 1ng (<1µl) of plasmid DNA (pDNA) was added to 40µl electrocompetent *E. coli* cells previously thawed on ice. Cells were then placed into a pre-chilled 1mm spaced electroporation cuvette (Invitrogen, UK) and pulsed using a BioRad MicroPulser<sup>™</sup> (BioRad, UK) at 1.8kV, 3-5ms. The cells were resuspended in 1ml SOC medium and incubated for 1h at 37°C with shaking at 275rpm (Innova<sup>®</sup>40, New Brunswick Scientific, USA) to aerate the culture. 10-100µl of the resuspended sample was plated on LB agar plates with the appropriate antibiotic. pCR<sup>®</sup>2.1 vector samples were spread onto LB agar plates containing 50µg/ml Kanamycin (*Kan*), 1mM IPTG, and 40µg/ml X-gal (Melford, UK) to facilitate blue/white selection.

TOP10F':	F' [ <i>lacI<sup>q</sup></i> Tn10 (Tet <sup>R</sup> )] <i>mcrA</i> Δ( <i>mrr-hsdRMS-mcrBC</i> ) φ80/ <i>lacZ</i> ΔM15 Δ <i>lacX74</i> <i>recA1</i> <i>araD139</i> Δ( <i>ara-leu</i> )7697 <i>galU galK rpsL</i> (Str <sup>R</sup> ) <i>endA1 nupG</i>
KRX:	[F', <i>traD36</i> , Δ <i>ompP</i> , <i>proA+B+</i> , <i>lacI<sup>q</sup></i> , Δ( <i>lacZ</i> )M15] Δ <i>ompT</i> , <i>endA1</i> , <i>recA1</i> , <i>gyrA96</i> (Nal(r)), <i>thi-1</i> , <i>hsdR17</i> (r(k)-, m(k)+), <i>e14-</i> (McrA-), <i>relA1</i> , <i>supE44</i> , Δ( <i>lac-proAB</i> ), Δ( <i>rhaBAD</i> )):T7 RNA polymerase

**Table 2.1**

*E.coli* genotypes used for cloning and expression studies.

### 2.2.8 Preparation of electrocompetent *E.coli*

The competent *E.coli* strains used in this study were prepared using the following method: 50ml sterile YENB medium was inoculated with a single colony of the desired *E.coli* strain and grown overnight at 37°C at 275rpm. 4L YENB medium (4x 1L, each in a 2L baffled shake flask) was inoculated each with 40ml of the overnight culture. The cultures were incubated at 37°C at 275rpm until an OD<sub>600</sub> of 1 was reached. The cultures were then



cooled on ice for 30min before centrifuging for 10min at 4,000xg. The supernatant was discarded and the cell pellet gently resuspended in 200ml ice-cold sterile Milli-Q water. The solution was centrifuged as before, and the supernatant discarded, thus removing residual salts from the sample. This step was repeated twice more and the pellet then resuspended in 100ml 10% (v/v) ice-cold glycerol in water and centrifuged as before. The supernatant was discarded and the cell pellet finally resuspended in 1ml 10% (v/v) ice-cold glycerol in water. The resuspended cells were dispensed into 40µl aliquots into pre-chilled (-70°C) 0.5ml tubes and stored at -70°C until required. The competency of the cells was measured by electroporating 1ng plasmid DNA and then plating a dilution of the resuspension onto LB plates with the appropriate antibiotic. Colonies were counted to estimate cell competency.

### **2.2.9 Colony screening**

Individual colonies were picked using a sterile toothpick and gridded onto an LB agar plate, with appropriate antibiotic, to assign each colony a clone number. The remaining cells attached to the toothpick were resuspended in 20µl sterile water and lysed by heating to 100°C for 5min, releasing the pDNA for use as template in a screening PCR. 50µl screening PCRs contained 1X MAXA *Taq* PCR Mastermix (GeneSys Ltd., UK), 25pmol each of vector-specific forward primer and vector-specific reverse primer, and 2µl lysed colony screen pDNA template, and was cycled on an Veriti® Thermal Cycler at 95°C – 3min, then 30 cycles (95°C – 10s, 55°C – 10s, 72°C – 1min/kb), 72°C – 5min, 4°C hold. 5µl PCR product was visualised on 1.5% (w/v) agarose gel stained with ethidium bromide to confirm amplification of the expected size fragment.

- M13\_Universal/M13\_Reverse primers (Appendix i.iii) were used with the pCR®2.1
- T7\_Promoter/T7\_Terminator primers (Appendix i.iii) were used with pET® vectors.

### **2.2.10 DNA sequencing**

#### *Template preparation:*

The PCR DNA fragment or plasmid DNA concentrations were estimated by agarose gel electrophoresis. PCR products were treated following the ExoSAP protocol before sequencing.

#### *ExoSAP reaction:*

1µl of the ExoSAP enzyme mix (5µl Exonuclease I (20U/µl, NEB) and 15µl Shrimp Alkaline

Phosphatase (1U/μl, NEB) was added to 50μl PCR product and incubated at 37°C for 60min. The enzymes were then inactivated by heating to 80°C for 15min. The PCR DNA template was then ready to be used directly in a sequencing reaction.

*Sequencing reaction:*

A 20μl reaction contained 1μl BigDye<sup>®</sup> Terminator V3.1 (ABI, UK), 7μl sequencing buffer, 3.2pmol primer, and 100-500ng DNA template, and was cycled on an ABI Veriti<sup>®</sup> Thermal Cycler at 96°C – 1min then 35 cycles (96°C – 10s, 50°C – 5s, 60°C – 4min), 70°C – 20min, 4°C hold.

*Sequencing reaction purification:*

Sequencing reactions were further processed to remove unincorporated dye terminators and reagents that may inhibit downstream applications. 5μl 125mM EDTA, (pH8.0 @ 25°C) was added directly to each sequencing reaction, followed by 2 volumes of 100% (v/v) ethanol. The reactions were incubated at -20°C for 30min before microfuging at 13,000xg for 10min. The supernatant was discarded and the DNA pellet was washed with 100μl 70% (v/v) ethanol in water, microfuged as before and the supernatant discarded. The DNA was then air dried for 10min to remove residual ethanol. The reactions were resuspended in 10μl Hi-Di<sup>™</sup> formamide (ABI, UK) and run on an Prism 3100XL (ABI, UK) Genetic Analyser. Sequencing data were analysed using ABI Sequence Scanner v1.0.

## **2.2.11 Gene-walking**

A gene-walking technique was employed to identify unknown DNA sequences flanking a region of known DNA sequence, i.e. the fragment identified from the *polA* screening method above.

*Genomic DNA digestion:*

gDNA was digested in 12 separate reactions, each containing a 6-cutter restriction endonuclease (RE) (NEB, UK). A 50μl digest reaction contained 1X CutSmart Buffer (NEB, UK), 100ng gDNA, 10U RE and 1X BSA when required. Reactions were incubated for 3h in a waterbath at the recommended reaction temperature for each RE and then heated to 85°C for 20min to denature the RE.

*Gene-walking primer design:*

Gene-specific gene-walking primer pairs (Appendix i.iv) were designed to target the sequenced DNA fragment, obtained using the *polA* screening PCR. Two primer pairs were

designed at the 5' and 3' ends of the DNA fragment, directing amplification towards the unknown DNA sequence required. This facilitated inverse PCR to improve the specificity of the reaction.

#### *Genomic DNA self-ligation reactions:*

gDNA fragments, digested as described above, were ligated to aid circularisation of small fragments, which thereby created a DNA template library for screening by PCR. 50ng of digested gDNA was ligated using 12.5U T4 DNA ligase (GeneSys Biotech Ltd., UK) with 1X T4 DNA ligase buffer (GeneSys Biotech Ltd., UK) in a 100µl reaction overnight at 16°C. The enzyme was then incubated at 85°C for 20min to denature the RE.

#### *Inverse PCR:*

Each of the 12 self-ligated DNA template libraries were used as template for a 50µl PCR containing 1x MAXA Taq Mastermix (GeneSys Biotech Ltd., UK), 25pmol of each inner gene-walking primer, and 1ng DNA template library, and were cycled on an ABI Veriti® Thermal Cycler at 95°C – 4min then 45 cycles (95°C – 10s, 55°C – 10s, 72°C – 5min), 72°C – 5min, 4°C hold. 2µl of a 1/100 dilution of the PCR was used as template in a 'nested Gene-Walking PCR' reaction.

#### *Nested gene-walking PCR:*

This additional PCR step was used to increase the specificity of template amplification. Reactions were carried out following the inverse PCR protocol but using the outer 'nested' primer pairs; 2µl of a 1/100 dilution of the inverse PCR product was used as DNA template, and an additional 10min was added to the final extension step at 72°C to ensure dA-tailing of the final DNA product to facilitate TA cloning. 5µl PCR product was visualised on a 1.5% (w/v) agarose gel stained with ethidium bromide to confirm amplification.

### **2.2.12 Sequence alignments**

DNA sequences were submitted to the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (nucleotide: BLASTN) (Altschul et al., 1990) and the translated amino acid sequences were submitted to protein BLAST (BLASTP), to compare with known sequences in the database. DNA sequences with high similarity to *polA* genes identified those sequences for further gene-walking, if required.

## 2.3 Methods: Molecular Cloning

### 2.3.1 Directional cloning primer design

Amino acid sequence alignments defined the positioning of the cloning primers to ensure the *polA* gene was cloned in the correct reading frame with the correct stop codon. Primers incorporated a unique 6-cutter RE site to facilitate directional cloning into pET<sup>®</sup> vectors.

#### *N-terminal primer design:*

The forward cloning primer contained 6 random nucleotides (N) followed by an *Nde* I RE site, incorporating the ATG start codon of the *polA* gene sequence, followed by 20-30 bases of gene specific sequence.

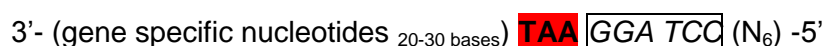


*Nde* I (CATATG) was used for the majority of clones at the N-terminal encoding region of the *polA* gene to position the ATG start codon for the correct open reading frame. pET24a+HIS was suitable for these reactions and is detailed on page 40. Where an internal *Nde* I site was found within a *polA* gene sequence, the alternative pET24d+HIS vector was used. This vector provided an alternative *Nco* I restriction site (CCATGG). The pET24d+HIS vector would not position the *polA* gene in the correct reading frame using *Nco* I and so an artificial DNA sequence 'AA' was added immediately after the 'G' of the ATG start codon to introduce an additional glycine (Gly/G) residue at the N-terminus of the Pol I protein. This extra amino acid was not expected to alter enzyme activity.



#### *C-terminal primer design:*

The reverse primer contained 20-30 bases of specific gene sequence followed by the *polA* stop codon (TAA), followed by either a *Bam*H I (G/GATCC) or *Sal* I (G/TCGAC) RE site and six random nucleotides (N) to allow effective cleavage by the RE.



### 2.3.2 High Fidelity Cloning PCR

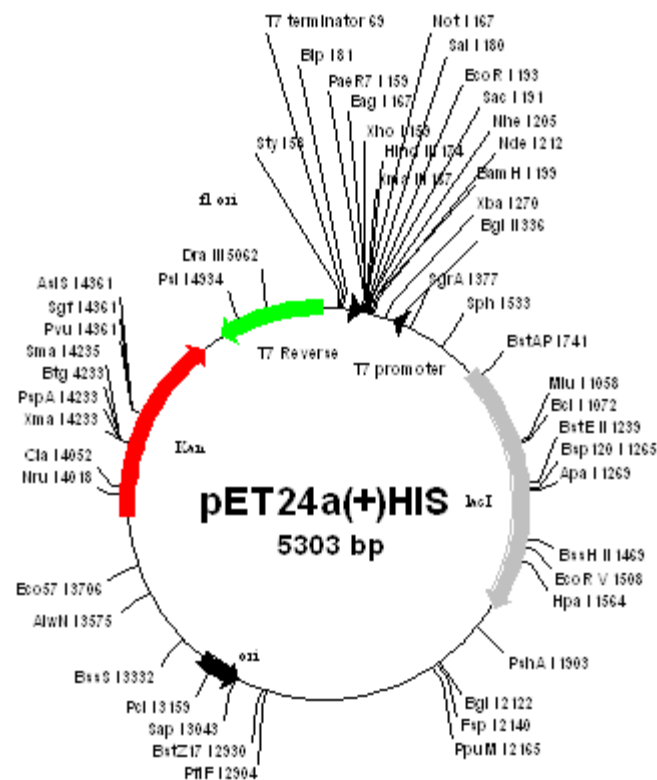
A high fidelity DNA polymerase enzyme was used to amplify fragments to be used for gene cloning. Phusion<sup>®</sup> DNA polymerase (Fisher Scientific, UK) provides an error rate 50-fold lower than conventional *T.aq* DNA Pol I, used for routine PCR. A 100µl high fidelity PCR contained 1X Phusion<sup>®</sup> HF Buffer (Fisher Scientific, UK), 0.4mM dNTPs (GeneSys Biotech Ltd., UK), 50pmol forward and reverse cloning primers, 10ng gDNA and 2U Phusion<sup>®</sup> DNA Polymerase. Reactions were cycled on an ABI Veriti<sup>®</sup> Thermal Cycler at 98°C – 30sec then 25 cycles (98°C – 5s, 58°C – 10s, 72°C – 30s/kb). 5µl PCR product was visualised on an agarose gel stained with ethidium bromide to confirm amplification. Fragments of the correct size were purified using PureLink PCR purification kits (Life Technologies, UK) following the manufacturer's instructions.

### 2.3.3 Restriction endonuclease digest

Restriction endonucleases (NEB, UK) were used to digest the purified DNA fragments and the appropriate pET<sup>®</sup> cloning vector to facilitate directional cloning. 100µl digest reactions contained 1X CutSmart Buffer (NEB, UK), 20U RE, and 1µg DNA, and were incubated at 37°C for 1h. An aliquot of the digested pET<sup>®</sup> vector was visualised on an agarose gel to confirm the dsDNA circular plasmid had been linearised to a single band. After confirmation, 10U Antarctic Phosphatase (AP) (NEB, UK) and 1X Antarctic Phosphatase Buffer (NEB, UK) were added directly to the digest and incubated at 37°C for 1h. The reaction removed the 5' phosphate group on the vector DNA, preventing self-ligation in downstream applications. Digested PCR and vector DNA products were purified using PureLink PCR clean-up kits (Life Technologies, UK), and an aliquot was run out on an agarose gel to estimate DNA concentration.

### 2.3.4 Ligations

Digested PCR product and pET<sup>®</sup> vector (Figure 2.2) were incubated in the presence of 10X T4 DNA ligase Buffer (GeneSys Biotech Ltd, UK) and 12.5U T4 DNA ligase (GeneSys Biotech Ltd., UK) in a 1:4 molar ratio of vector to insert. 10µl ligation reactions were incubated overnight at 16°C and then heat treated at 70°C for 20min to denature the T4 DNA ligase enzyme. The ligated DNA was ethanol precipitated and resuspended in Milli-Q water ready for electroporation.



**Figure 2.2**

Key features of the pET<sup>®</sup> vector used for directional cloning in this study. The pET24a+ vector has been further modified by GeneSys Biotech Ltd. to incorporate an N-terminal histidine (6X-His) tag to aid recombinant protein purification (see Appendix v.i for a detailed description).

### 2.3.5 Plasmid mini-prep

Recombinant colonies, determined by the colony screening PCR method, were inoculated into 10ml LB media containing 50µg/ml *Kan* and incubated at 37°C with aeration at 275rpm overnight. Cultures were then centrifuged at 5,000xg for 10min and the supernatant discarded. pDNA was then extracted from the cell pellet using a PureLink Plasmid purification kit (Life Technologies, UK) following the manufacturer's instructions. Purified pDNA was eluted in Milli-Q water and an aliquot visualised on a 1% (w/v) agarose gel, stained with ethidium bromide, to estimate the concentration of the DNA.

### 2.3.6 Site-directed mutagenesis

An overlap extension PCR method was employed to introduce single or multiple nucleotide mutations into the *polA* gene sequence (Appendix i.vi). This overlap method could also be employed for the creation of chimeric enzymes (Appendix iii). Mutagenic primers were designed to introduce the desired mutation (or chimera) to both strands of the amplified PCR fragment (Figure 2.3).

```
Tin_WT sequence   .....CGAAGTGGTGGCTATTGACCTTGAAAGTGATACGAAAGAC.....
Tin_3-5exoKO_U    5'-CGAAGTGGTGGCTATTGCTCTTGCTAGTGATACGAAAGAC-3'
Tin_3-5exoKO_L    5'-GTCTTTCGTATCACTAGCAAGAGCAATAGCCACCACTTCG-3'
```

**Figure 2.3.**

An example of mutagenic primers employed in overlap extension PCR to mutate the blue highlighted bases to those in red. A detailed explanation of the primers is given in Chapter 6.

Two PCRs were required for each desired mutation. The first PCR employed the vector-specific forward primer (T7 Promoter) with the mutagenic lower primer. The second PCR employed the vector-specific reverse primer (T7 Terminator) with the mutagenic upper primer. A 50µl high fidelity PCR contained 1X Phusion® HF Buffer, 0.2mM dNTPs, 25pmol of each forward and reverse primers, 10ng pDNA and 2U Phusion® DNA Polymerase. Reactions were cycled on an Veriti® Thermal Cycler at 98°C for 30sec then 15 cycles (98°C – 5s, 58°C – 10s, 72°C – 30s/kb), 72°C – 5min, 4°C hold. 10U *Dpn* I RE (NEB, UK) was added to each PCR to remove the parental pDNA from the reaction. 2µl of a 1/100 dilution

of each PCR was used as DNA template for a final round of PCR to generate the mutant fragment. A 100µl high fidelity PCR contained 1X Phusion® HF Buffer, 0.2mM dNTPs, 25pmol T7 Promoter and T7 Terminator primers, 2µl (1/100 dilution) *Dpn* I treated PCR product and 2U Phusion® DNA Polymerase. Reactions were cycled on an ABI Veriti® Thermal (98°C – 30sec then 25 cycles (98°C – 5s, 58°C – 10s, 72°C – 30s/kb), 72°C – 5min, 4°C hold. 5µl PCR product was visualised on an agarose gel, stained with ethidium bromide, to confirm amplification. Fragments of the correct size were purified using PureLink PCR purification kits following manufacturer's instructions and directionally cloned into a pET® vector.

### **2.3.7 -70°C Freezer stocks**

Recombinant clones that had been sequence verified were streaked out onto an LB agar plate with appropriate antibiotic and incubated overnight at 37°C. The cells were transferred from the plate to a 2ml Cryotube (Fisher Scientific, UK) containing 1ml LB and 0.5ml 50% (v/v) glycerol in water, and the resuspension was stored at -70°C until required. To revive the strain, a sterile loop was used to scratch the frozen surface of the stock and streaked onto a fresh LB agar plate with appropriate antibiotic, which was then placed at 37°C overnight to revive the clone.

## **2.4 Methods: Protein expression**

### **2.4.1 Small-scale protein expression**

Recombinant clones (pET\_*polA* construct) were revived on LB agar plates containing 50µg/ml kanamycin and 34µg/ml chloramphenicol and incubated at 37°C overnight. A single colony was used to inoculate 2.5ml LB containing antibiotics as before, which was then incubated at 37°C with aeration at 275rpm overnight. 100µl culture was inoculated into 10ml TB containing 50µg/ml kanamycin and grown at 37°C with aeration at 275rpm until an OD<sub>600</sub> of 0.8 was reached. An identical sample was grown alongside to be used as an un-induced control. The temperature of incubation was lowered to 24°C and protein expression was induced with 0.1% (w/v) L-rhamnose (Rose Scientific, UK) and 1mM IPTG (Melford, UK). Cultures were incubated overnight (22h) and the OD<sub>600</sub> recorded. The induced cells were harvested by centrifugation at 5,000xg for 10min, the supernatant discarded, and the cell pellet frozen at -70°C until required.



### **2.4.2 Large-scale protein expression**

Recombinant clones (pET\_*polA* construct) were revived and a 2.5ml overnight LB culture was set up as described in 2.4.1. 1ml of the culture was inoculated into 100ml TB containing 50µg/ml kanamycin and induced following method 2.4.1. A 10ml aliquot was removed prior to induction, centrifuged at 5,000xg and the cell pellet frozen at -70°C for use as an un-induced control sample.

### **2.4.3 96-well plate protein expression**

To enable the direct comparison of multiple clones, a 96 deep-well plate was used to express a large number of clones under identical growth conditions. The required pET\_*polA* construct was inoculated from a single revived colony into 1ml LB containing 50µg/ml kanamycin and 34µg/ml chloramphenicol in a 2ml well of a 96 deep-well plate (Fisher Scientific, UK). The plate was sealed with a breathable self-adhesive plate seal (Fisher Scientific, UK), placed at an angle of 45°C and incubated at 37°C with aeration at 275rpm overnight (22h). The OD<sub>600</sub> was recorded, to confirm each colony had grown to the same density, and 10µl sample used to inoculate 1ml TB, containing 50µg/ml kanamycin, pre-aliquoted into a fresh 96 deep-well plate. An identical sample was cultured to be used as an un-induced control. The plate was sealed and incubated as before. When the cultures reached an OD<sub>600</sub> of 0.8, the temperature of incubation was lowered to 24°C and protein expression was induced following the method 2.4.1. The overnight induced cells were harvested by centrifugation at 2,750xg for 15min using an Eppendorf centrifuge 5804 (USA). The supernatant was discarded and the plate, containing pelleted cells, stored at -70°C until required.

### **2.4.4 Protein sample preparation**

Induced cell pellets, obtained from either the small scale or 96 well plate cultures, were thawed on ice and resuspended in 1ml cell lysis buffer. Large scale expression pellets (from a 100ml culture) were resuspended in 10ml Purification Buffer A. Cell suspensions were vortexed briefly to mix and sonicated on ice at 40W to lyse the cells and release the induced soluble protein. Large scale cultures were sonicated for 2x 1min, small scale for 30s, and 96 well scale for 10s. Sonication was achieved using an ultrasonic homogeniser (4710 series) with 5mm Ø probe (Cole-Palmer instruments, USA). Samples were heat treated at 65°C in a waterbath for 60min to denature *E.coli* host proteins. The cell debris, and denatured and insoluble protein material, were separated from the soluble proteins by centrifugation at

10,000xg for 15min. The supernatant was transferred to a fresh tube and used as the soluble protein fraction. The remaining cell pellet was washed with an equal volume of cell lysis buffer, pelleted as before and then mixed with an equal volume of cell lysis buffer. This fraction was used to identify insoluble proteins. An aliquot of each fraction was mixed with an equal volume of protein loading buffer and denatured in boiling water for 5min. Samples were loaded directly onto a pre-cast 12% SDS-PAGE gel for protein analysis.

#### **2.4.5 SDS Polyacrylamide Gel Electrophoresis (SDS-PAGE)**

Protein samples were loaded into pre-cast 12% SDS-PAGE gels (GeneSys Biotech Ltd., UK) in Mini-Protean tanks (BioRad, UK) with 10µl PageRuler protein ladder (Fisher Scientific, UK) to estimate protein size in kDa. Gels were electrophoresed at 75V in SDS-PAGE protein running buffer to migrate the proteins through the stacking gel. Upon reaching the resolving gel, the voltage was increased to 175V until the dye in the protein loading buffer reached the end of the gel. Gels were transferred to SDS-PAGE de-stain buffer to remove residual SDS from the gel surface, and then stained with SDS-PAGE stain for 30min. The de-stain buffer was used to remove background dye from the gel, revealing expressed proteins.

### **2.5 Methods: Protein purification**

The ÄKTA purification system used in this study was purchased from GE Healthcare (UK). All buffers were made using Milli-Q water and autoclaved at 120°C for 30min. Soluble Pol I protein was purified using a three column approach: (i) Metal-chelating affinity chromatography, (ii) Heparin affinity chromatography and a final polishing step using (iii) Ion-exchange chromatography.

#### **2.5.1 Metal-chelating affinity column chromatography**

##### *Sample preparation:*

The heat-treated soluble protein sample, obtained from the large-scale induction culture, was increased to a final volume of 20ml using Purification Buffer A. A 100µl aliquot was removed to be used as an example of the load, for purification control.

*Column equilibration:*

The 1ml Ni-NTA HisPur™ column (Fisher Scientific, UK) was washed with 5 column volumes (CV) of Milli-Q water to remove the column storage buffer, then equilibrated with 5CV of Purification Buffer A. All buffers were run at 1ml/min.

*Sample purification:*

The 20ml protein sample was loaded onto the nickel charged column and the flow-through retained. The column was then washed with 5CV of Purification Buffer A to remove remaining un-bound proteins. The bound protein sample was eluted with a 0-100% gradient (5-400mM Imidazole) of Purification Buffer B over 10CV, collecting 0.5ml fractions. Fractions containing the eluted protein were analysed using SDS-PAGE, loading the control 'non-purified' and 'flow-through' samples alongside.

## **2.5.2 Heparin column chromatography**

*Sample preparation:*

Fractions identified to contain a protein band at the expected molecular weight (60-70kDa for DNA Pol I proteins) were pooled and transferred to sterile dialysis tubing and dialysed sequentially against 2x 500ml Purification Buffer C to reduce the NaCl concentration. The sample volume was increased to 20ml by the manual addition of Purification Buffer C and a 100µl sample taken as a purification control sample.

*Column equilibration:*

A 5ml HiTrap™ HP column (GE Healthcare, UK) was washed with 5CV of Milli-Q water (to remove the column storage buffer), then equilibrated with 5CV of Purification Buffer C, at 1ml/min.

*Sample purification*

Pooled and dialysed fractions from the Ni-NTA HisPur™ column were loaded onto the column and the flow-through collected. The column was washed with 5CV Purification Buffer C and the bound protein eluted using a 0-100% gradient (0.05M-1M NaCl) of Purification Buffer D, over 10CV, collecting 0.5ml fractions, at 1ml/min. Successful protein purification was confirmed using SDS-PAGE as before.

### **2.5.3 Ion exchange column chromatography**

#### *Sample preparation*

To purify the enzyme further, pooled fractions from the previous purification step were sequentially dialysed against 2x 500ml Purification Buffer C to reduce the NaCl concentration and a 100µl sample taken as a purification control sample.

#### *Column equilibration:*

A 1ml ResourceQ anion exchange column (GE Healthcare, UK) was washed with 5CV of Milli-Q water to remove the column storage buffer, and then equilibrated with 5CV of Purification Buffer C, at 1ml/min. Where the isoelectric point (pI) of the protein of interest was greater than the buffer pH (pH 8.0), a 1ml ResourceS (GE Healthcare, UK) cation exchange column was used.

#### *Sample purification:*

The dialysed and pooled fractions purified on the Heparin column were loaded onto the ion-exchange column and the flow-through collected. The column was washed with 5CV Purification Buffer C and the protein eluted using a 0-100% gradient (0.05M-1M NaCl) of Purification Buffer D, over 10CV, collecting 0.5ml fractions, at 1ml/min. Successful protein purification was confirmed using SDS-PAGE as before. Desired fractions were pooled and sequentially dialysed against 2x 250ml Polymerase Storage Buffer. A 100µl aliquot was transferred to a 0.5ml tube and frozen at -20°C until required for protein concentration and thermal melt analysis. The non-ionic detergents Triton X-100 and IGEPAL<sup>®</sup> CA-630 were each added to the remaining 1-2ml protein sample to a final concentration of 0.5% (v/v), with 1mM DTT further included, for long-term storage at -20°C.

### **2.5.4 Protein concentration**

The Qubit<sup>®</sup> protein concentration kit (Life Technologies, UK) was used to calculate the purified protein yield from each Pol I sample preparation. A dilution of purified sample (in a storage buffer minus detergent) was measured using the kit, following the manufacturer's instructions. A known concentration of BSA, supplied in the kit, was used to create a standard curve against which the Pol I samples could be quantified. Samples were loaded into a 96 well flat-bottom plate (Fisher Scientific, UK) and the fluorescence analysed using a Synergy HT fluorescent plate scanner (Bio-Tek, UK) reading excitation at 485nm and emission at 560nm. Sample concentrations were calculated as mg/ml enzyme and the total yield from the 100ml TB culture recorded as mg/L culture for comparison.

## 2.6 Methods: Characterisation

### 2.6.1 Nuclease assay

100ng DNA Pol I was added to 1X iBuffer (OptiGene Ltd., UK) and 500ng of a variety of nucleic-acid templates including: (i) Lambda DNA digested with *EcoRI/HindIII*, (ii) pET<sup>®</sup> vector DNA (Novagen, UK), (iii) Lambda DNA (Fermentas, UK), and (iv) MS2 RNA (Fermentas, UK). The 20µl reaction was incubated for 3h at 37°C and then electrophoresed on a 0.7% (w/v) agarose gel. Samples were confirmed to be nuclease-free if the nucleic-acid stained bands matched those of control reactions containing no enzyme.

### 2.6.2 Genomic DNA contamination assay

Purified DNA Pol I enzyme was assessed for contaminating nucleic-acids that may have carried through from the column purification procedure. Enzyme was heat treated at 98°C for 30min to denature the DNA Pol I. Samples were diluted to 10ng/µl in TE buffer and 10µl (100ng) used in a qPCR. 23S rDNA target primers (GeneSys Biotech Ltd., UK) were used to screen for contaminating *E.coli* nucleic-acids within the sample. A serial dilution of *E.coli* (strain W1485) gDNA (GeneSys Biotech Ltd., UK) was used as a comparison for amplification. A 50µl qPCR contained 1x *T.aq* DNA Pol I-based qPCR Mastermix (containing SYBR<sup>®</sup> Green I), 15pmol 23S rDNA primer mix, 10µl gDNA template dilution or 10µl heat treated enzyme sample (@10ng/µl), and DNase-free water (GeneSys Biotech Ltd., UK). Reactions were run on an Light Cycler<sup>®</sup> LC-480 (Roche, USA) at 95°C for 3min then 55 cycles of: 95°C for 5s, 60°C for 15s. A melt profile was run consisting of: 60°C-95°C ramping at 0.04°C/s to confirm the reaction products.

### 2.6.3 Thermal shift assay

The thermostability of the DNA Pol I enzymes were compared using the fluorescent protein dye SYPRO<sup>®</sup>-orange (Life Technologies, UK). A 20µl reaction contained 1X iBuffer (including 5mM MgSO<sub>4</sub>, but minus detergent), 10X SYPRO<sup>®</sup>-orange dye, and 1µg protein sample. Reactions were mixed and transferred to optically clear Genie<sup>®</sup>II tubes (OptiGene Ltd., UK) and analysed on a Genie<sup>®</sup>II (OptiGene Ltd., UK) real-time fluorescence detection instrument, with heating from 35-105°C, ramping at 0.05°C/s, taking readings every 5s. A derivative plot of the thermal melt data, displayed as an increase in fluorescence with increasing temperature, gave a threshold figure ( $T_m$ ) for the maximum rate of change, with which comparisons between enzymes could be directly drawn.

#### 2.6.4 Strand-displacing primer extension assay

A gel-based M13mp18 ssDNA primer extension assay was used to assess the novel DNA Pol I enzymes ability to strand-displace DNA. Two primers were designed for the assay: a -47 primer; and a Blocking primer, 5kb downstream of the -47primer, containing a phosphorothioate base at the n-1 position at the 3'end.

##### *Primer annealing:*

A 25µl primer annealing reaction contained 0.5µg M13mp18 ssDNA (GeneSys Biotech Ltd., UK), 40pM -47 primer, 40pmol Blocking primer, and 1X iBuffer (including 5mM MgSO<sub>4</sub>). The reaction was incubated at 90°C for 2min then cooled to room temperature over 20min to ensure successful annealing of the primers to the template DNA.

##### *Primer extension:*

The 25µl annealed primer/template mix was added to a 25µl reaction mix containing 66.6ng DNA Pol I enzyme (equivalent to 8U *Bst* DNA Pol I), 0.4mM dNTP mix, and 1X iBuffer (including 5mM MgSO<sub>4</sub>). The 50µl reaction mix was set up on a cool block to prevent activity. Reactions were incubated at 65°C, 75°C, and 85°C for 30min. Reactions were then immediately cooled and 20µl reaction product mixed with 5µl gel loading solution containing 10mM EDTA to stop the reaction. The 25µl mix was electrophoresed on a 0.7% agarose gel compared to a no-enzyme control (NEC) sample. 5'-3' DNA Pol I activity without strand-displacement activity was shown by the amplification of a 5kb DNA fragment, stalled at the blocking primer. The presence of DNA larger than 5kb indicated strand-displacement activity i.e. the ability to displace the blocking primer.

#### 2.6.5 Loop-mediated isothermal DNA amplification

This assay demonstrates the use of the DNA Pol I enzymes in an isothermal strand-displacing DNA amplification reaction. All LAMP primers were designed using LAMP Designer<sup>TM</sup> software (Premier Biosoft International, USA). 25µl reactions contained 1x iBuffer (including 5mM MgSO<sub>4</sub>) (OptiGene Ltd., UK) or 1x ThermoPol Buffer (including 5mM MgSO<sub>4</sub>) (NEB, UK), 1M Betaine, 0.4mM dNTP mix, 10ng *Aeropyrum pernix* (*A.pe*) gDNA (GeneSys Biotech Ltd., UK), 1X LAMP Primer mix (0.8µM *Ape\_FIP*, 0.8µM *Ape\_BIP*, 0.2µM *Ape\_F3* 0.2µM *Ape\_B3*, 0.4µM *Ape\_LoopF*, 0.4µM *Ape\_LoopB*), 0.125X EvaGreen<sup>®</sup> (Biotium, USA), and 8U *B.st* LF DNA Pol I (NEB, UK) or 1µl purified LF DNA Pol I enzyme dilution. Reactions were run for 60min at 65°C on an LC-480 (Roche, USA) or Genie<sup>®</sup>II

(OptiGene Ltd, UK) fluorescent detection instrument for real-time amplification detection. Non-fluorescent reactions (minus EvaGreen®) were incubated in a 65°C waterbath for 60min, and 5µl visualised on a 1% (w/v) agarose gel to confirm amplification. With reactions generating >25µg amplicon (Notomi, 2000), reaction tubes were opened in a separate laboratory to prevent contamination with the specific *A.pe* product.

#### **2.6.6 Reverse transcription LAMP**

This method demonstrates the ability of the DNA Pol I enzyme to carry out RT-LAMP. 25µl reactions contained 1x iBuffer (including 5mM MgSO<sub>4</sub> or 5mM MnCl<sub>2</sub>) (OptiGene Ltd., UK), 1M Betaine, 0.4mM dNTP mix, 10ng MS2 RNA (Fermentas, UK), 1X RT-LAMP Primer mix (0.8µM MS2\_FIP, 0.8µM MS2\_BIP, 0.2µM MS2\_F3 0.2µM MS2\_B3, 0.4µM MS2\_LoopF, 0.4µM MS2\_LoopB), 0.125X EvaGreen®, and 250ng purified LF DNA Pol I. Identical reactions were run with the addition of 0.25U AMV-reverse transcriptase (GeneSys Biotech Ltd., UK) as a control. Reactions were run for 60min at 65°C on an LC-480 (Roche, USA).

#### **2.6.7 Shimadzu analyser of amplification reactions**

5µl LAMP-derived amplicons were analysed using the MultiNA analyser (Shimadzu Biotech, Japan), following the manufacturer's protocol to assess the fragment sizes and yield of the amplified DNA.

#### **2.6.8 Buffer optimisation**

The 10X iBuffer (OptiGene Ltd., UK) was re-optimised for use in LAMP. The MgSO<sub>4</sub>, KCl and Tris-HCl concentrations were assessed. A 10X iBuffer, minus the component under test, was prepared and a dilution of the component was then run in the buffer. Reactions were run following the LAMP method in 2.6.5, with 250ng LF DNA Pol I. Optimised components were defined as the fastest time to result with 250ng LF DNA Pol I at 65°C.

#### **2.6.9 Incorporation of dUTP**

The ability of each novel DNA Pol I to incorporate dUTP was assessed using the LAMP method. Reactions were set up following method 2.6.5 minus the dNTP mix. LAMP reactions containing 250ng DNA Pol I were run with a dilution of a dUTP containing nucleotide mix (0-100% dUTP in place of dTTP). A comparison of amplification times was used to assess the enzymes ability to incorporate dUTP at 65°C.

### 2.6.10 *B.st* LF DNA polymerase I unit assay

An M13mp18 ssDNA primer extension assay was used to compare the extension rate of the novel DNA Pol I enzymes to known concentrations of *B.st* LF DNA Pol I (NEB, UK), under the same experimental conditions.

#### *Primer annealing:*

A 140µl reaction containing 28µg (14pmol) M13mp18 ssDNA, 24pmol -47 primer, 1X iBuffer (including 5mM MgSO<sub>4</sub>) was heated to 90°C for 1min and then cooled to room temperature over 20min to ensure successful primer annealing to the template DNA.

#### *Primer extension:*

1.2µl primer/template mix was added to a reaction mix containing 1X iBuffer (including 5mM MgSO<sub>4</sub>), 0.4mM dNTP mix, and 0.125X EvaGreen<sup>®</sup>, made up to 24µl with DNase-free water. 1µl of a 1 in 2 serial dilution of DNA Pol I stocks was used in each reaction and compared to a dilution of *B.st* LF DNA Pol I (120,000units/mg) (NEB, UK). Reactions were set up on a cool block to prevent enzyme activity and run on an LC-480 (Roche, USA) monitoring fluorescence at 65°C for 30min. An increase in fluorescence over background identified an active enzyme, with rates of fluorescence compared to those of the *B.st* LF DNA Pol I control.

### 2.6.11 Polymerase Chain Reaction

Highly thermostable DNA Pol I were assessed for their ability to amplify DNA in a PCR. Purified Pol I was used to amplify DNA fragments of 250, 500, 750, 1000, 2000, 3000bp to compare the processivity of the different enzymes. Each 50µl PCR contained 1X PCR Buffer (GeneSys Biotech Ltd., UK), 0.4mM dNTPs (GeneSys Biotech Ltd., UK), 25pmol forward and reverse primers (see Appendix i.vii), 10ng Lambda DNA (Fermentas, UK) and a dilution of DNA Pol I. Reactions were cycled on an ABI Veriti<sup>®</sup> Thermal Cycler at 90°C – 10sec then 25 cycles (90°C – 5sec, 55°C – 10s, 72°C – 1m). 5µl PCR product was visualised on a 1% (w/v) agarose gel, stained with ethidium bromide, to confirm amplification.



### 3. IDENTIFICATION

#### 3.1 Introduction

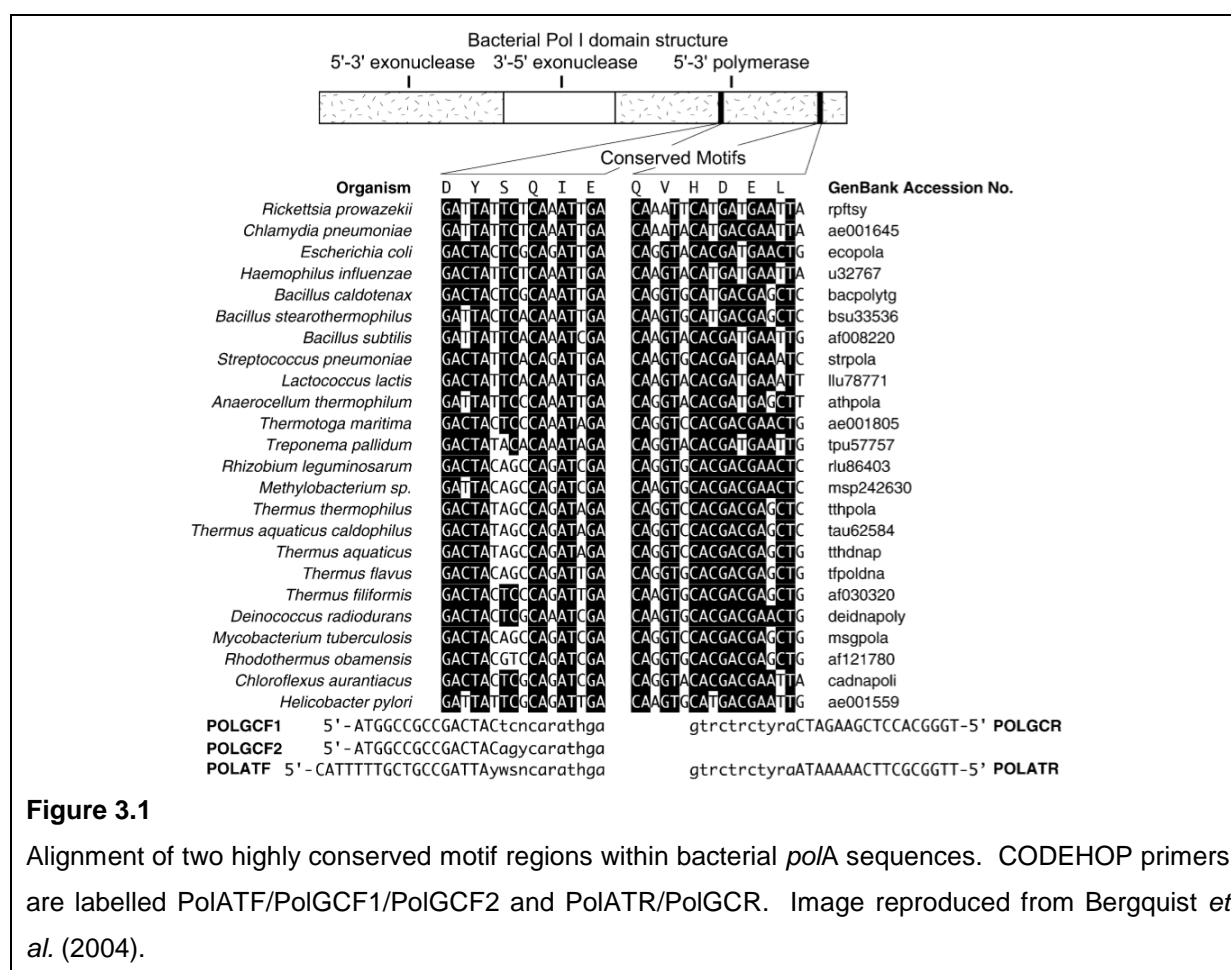
*B.st* Large Fragment DNA polymerase I, the current enzyme of choice for isothermal DNA amplification reactions, was first isolated by Stenesh and Roe (1972) from the thermophilic bacterium *Geobacillus stearothermophilus* (*B.st*), which grows between 45 and 75°C. *B.st* LF DNA Pol I is active at an optimal temperature of 65°C, and is inactivated after 15 min incubation at 75°C (NEB, UK).

First isolated in 1972, Kaboev *et al.* (1981) further purified the enzyme, as did Sellman *et al.* (1992). Alliota *et al.* (1996; and 1998) purified *B.st* DNA Pol I from *Geobacillus stearothermophilus* strain N3468 and confirmed the enzyme lacked 3'-5' exonuclease activity. Since its isolation, further thermophilic *Geobacillus* DNA polymerases have been characterised (Uemori *et al.*, 1993; Sandalli *et al.*, 2009; Caglayan *et al.*, 2011). A number of viral and bacterial DNA polymerases have been reported identifying strand-displacement activity (Blanco *et al.*, 1984), but these mesophilic enzymes are not thermostable enough to withstand the high temperatures required for high primer specificity for diagnostic nucleic acid amplification techniques. The moderately thermophilic *B.st* LF DNA Pol I has become the enzyme of choice, and is widely reported in publications where a strand-displacing, moderately thermostable enzyme is required. As such, *B.st* LF DNA Pol I is to be used as the gold-standard enzyme in this report.

*Geobacillus*-like DNA polymerases have been previously identified. Uemori *et al.* (1993) reported the successful cloning of *Bacillus caldotenax* DNA *polA* into *Escherichia coli* with no prior knowledge of the gene or protein sequence. Degenerate primers were designed using alignments of *polA* sequence from *Escherichia coli*, *Streptococcus pneumoniae*, *Thermus aquaticus* and *Thermus thermophilus*. The use of consensus-degenerate hybrid oligonucleotide primers (CODEHOP, Rose *et al.* 1998) have more recently been shown to be effective in the isolation of unknown sequences. Previous methods to isolate unknown family members by PCR have relied on either degenerate primers consisting of a pool of primers containing most or all of the possible nucleotide sequences encoding a conserved amino acid motif or consensus primers consisting of a single primer containing the most common nucleotide at each codon position within the motif. Although these strategies have been successful in isolating closely related sequences, they have generally failed when sequences were more distantly related or were in low copy number (Rose *et al.*, 2003).

CODEHOP primers contain a 3' degenerate core that is relatively short with only 3-4 conserved residues required. This allows for codon usage within an organism and prevents a low yield of individual primers that would otherwise be used up early in the reaction cycles. The 5' end of the primers contain a non-degenerate clamp that allows higher annealing temperatures in later cycles but does not add to the degeneracy of the primer (Rose *et al.*, 1998).

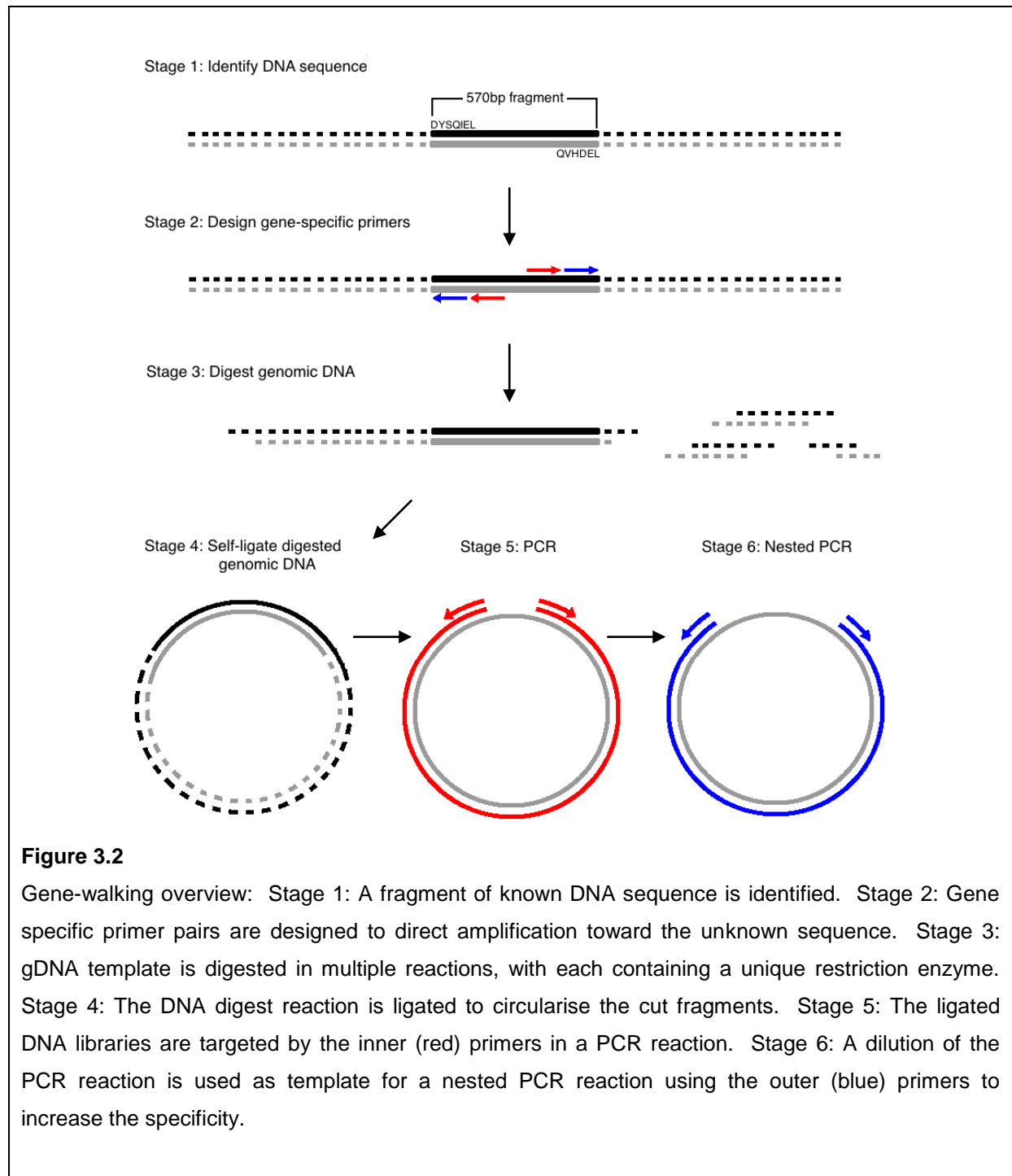
The isolation of novel Pol I enzymes from thermophilic organisms has also been reported by Bergquist *et al.* (2004). Conserved motif regions of known bacterial *polA* sequences were identified and degenerate 'screening' primers designed to amplify an internal portion of the gene by PCR. Bergquist *et al.* (2004) utilised CODEHOP primers to hybridise to the DNA coding for two highly conserved regions (motif 'DYSQIE' and motif 'QVHDEL') identified by aligning 24 different bacterial *polA* sequences (Figure 3.1). A Touch-Down PCR method was used where the high initial annealing temperature was decreased by 1°C per cycle. This cycling protocol increases the specificity of the PCR because only exact matches will anneal to the primers at the higher temperatures. A gradual reduction in temperature ensures specific primer annealing will take place.



**Figure 3.1**

Alignment of two highly conserved motif regions within bacterial *polA* sequences. CODEHOP primers are labelled PolATF/PolGCF1/PolGCF2 and PolATR/PolGCR. Image reproduced from Bergquist *et al.* (2004).

Triglia *et al.* (1989) identified a method known as inverse PCR to amplify DNA sequences that lie outside the boundaries of known sequences. The technique requires the inversion of the sequence of interest by circularisation and re-opening at a different site. The genome of the organism of interest is cleaved with a restriction site and later ligated to re-circularise the DNA into smaller fragments. These new, smaller genome fragments are then screened using known PCR primers to specifically amplify sequences up- and down-stream of the known fragment. Figure 3.2 describes an overview of the gene-walking method.



**Figure 3.2**

Gene-walking overview: Stage 1: A fragment of known DNA sequence is identified. Stage 2: Gene specific primer pairs are designed to direct amplification toward the unknown sequence. Stage 3: gDNA template is digested in multiple reactions, with each containing a unique restriction enzyme. Stage 4: The DNA digest reaction is ligated to circularise the cut fragments. Stage 5: The ligated DNA libraries are targeted by the inner (red) primers in a PCR reaction. Stage 6: A dilution of the PCR reaction is used as template for a nested PCR reaction using the outer (blue) primers to increase the specificity.

It will be important to type strain the organisms from which the newly identified *polA* genes are found. This can be achieved through a comparison of the unique 16S rRNA gene sequence identified from each organism. 16S rRNA genes are an essential component of the transcriptional machinery of prokaryotes and occur in at least one copy in a genome. The universality of the 16S rRNA genes enables them to be targeted for phylogenetic analysis and taxonomic classification (Woese, 1987). Initially, primers directed to the gene were designed by targeting conservative regions of 16S rDNA alignments to *E.coli* (Lane *et al.* 1985) but, over time, with the increasing number of known 16S rDNA sequences and polymorphisms available, primers targeting specific phyla were designed (Lane *et al.* 1991). A selection of these published primers are expected to allow effective type straining, and thus comparison of the organisms used in this study

### 3.1.1 Chapter overview

It is proposed that organisms isolated from temperatures close to the optimal LAMP operating temperature of 65°C will harbour enzymes stable and optimally active at those temperatures. A wide variety of thermophilic organisms were chosen for this study; some have genome sequences publically available from the NCBI database, whilst others are merely available from culture collections with little or no sequence data known. Further organisms have been isolated from soil sample collections (GeneSys Biotech Ltd., UK) that had been gathered over the last 30 years, with additional isolates sampled directly during this study from the Azores (Mid-Atlantic Ridge, Portugal). The use of 'Universal' 16S rRNA primers will enable PCR amplification of the 16S rRNA gene sequences of the organisms in this study. Taxonomic alignment will provide a greater understanding of the homology between strains and the novel enzymes that they may harbour. The results from the Uemori and Bergquist groups suggests CODEHOP style primers, based upon common genetic motif regions, are suitable to identify gene fragments from distantly related organisms and would therefore be suitable for this study.

This chapter details the isolation of thermophilic organisms and the subsequent extraction of genomic DNA to enable 16S rRNA gene alignment and taxonomic comparisons. Organisms without publically available genomic DNA sequences were screened using CODEHOP *polA* primers in a touch-down PCR. The inverse PCR gene-walking method, identified by Triglia *et al.* (1989), was then used to reveal the whole *polA* gene sequence for further study. Alignments of the Pol I amino acid sequences were anticipated to identify common *polA* regions and motifs suggesting potential enzymes suitable for isothermal reactions.

## 3.2 Results

### 3.2.1 Isolation of thermophilic strains

Thermophilic strains were obtained from a number of different sources for this study. The simplest route to obtaining a diverse collection was to purchase organisms from the DSMZ microorganism culture collection (Germany), selecting strains with growth temperature optima close to 65°C. The majority of samples arrived as lyophilised cell pellets but others as live, liquid cultures. Live cultures were incubated for a further 24h at their optimal growth temperature to encourage additional growth. Strains were also donated for this study. *Geobacillus kaustophilus* HT24A was kindly donated by Dr Takami (Microbial Genome Research Group, Japan Agency of Marine-Earth Science and Technology) and arrived as an LB stab. *Geobacillus caldovelox* was donated by Dr Clark (GeneSys Biotech Ltd., UK).

A large variety of soil samples, isolated from around the world, were available from GeneSys Biotech Ltd. to be screened for *Geobacillus* spores. Six samples were chosen for this work covering a wide geographical area (Table 3.1). Additionally, I was able to sample thermal features directly in the Azores (Extremophiles conference, Portugal, 2010) providing a variety of soil and thermal run-off water samples to screen for additional thermophilic organisms. Strains were revived following the methods described in Chapter 2.2.1 at a variety of temperatures, then sequentially plated to ensure the presence of a single strain. Table 3.2 provides information on the individual organisms selected from the DSMZ culture collection, and those kindly donated for this study.

Sample	Location/Type	Donated by	Revival temp °C
MeIA	Melbourne/Soil	GeneSys Biotech Ltd.	60°C
MeIB	Melbourne/Soil	GeneSys Biotech Ltd.	55°C
MeIC	Melbourne/Soil	GeneSys Biotech Ltd.	55°C
HK50b	Hong Kong/Soil	GeneSys Biotech Ltd.	55°C
SA1	South Africa/Soil	GeneSys Biotech Ltd.	50°C
Zim1	Zimbabwe/Soil	GeneSys Biotech Ltd.	55°C
Azores No.1-10	Azores/liquid 50-90°C pH2-6.0)	Self-sampled	No.8 @ 55°C
Azores No.11-18	Azores/Soil 45-99°C (pH2-5.5)	Self-sampled	No.13/15 @ 45°C

**Table 3.1**

Samples screened for thermophilic organisms. The revival temperature on LB plates, the sampling environment and location are reported.

Organism	Isolation	Growth characteristics	Reference
<i>Thermus aquaticus</i> YT-1 ATCC 25104, DSM 625 ( <i>T.aq</i> )	Yellowstone National Park Hot Springs (USA)	Aerobic, Gram-negative, non-motile, non-sporulating rods and filaments 0.5-0.8µm in diameter. Growth temperature 50-80°C, with optimum at 70-75°C, pH7.5-7.8. Strain YT-1 – 67.4% G+C DNA base composition.	Brock <i>et al.</i> 1969
<i>Thermotoga maritima</i> MSB8 JCM 10099, DSM 3109 ( <i>T.ma</i> )	Vulcano island	Anaerobic, rod-shaped, fermentative, extremely thermophilic and grows between 55 and 90°C with an optimum of around 80°C. pH5-9 with an optimum at pH 6.5. Rods are 1.5-11µm in diameter.	Stetter <i>et al.</i> 1986
<i>Geobacillus stearothermophilus</i> ( <i>B.st</i> )	New England Biolabs isolate N3468	Thermophilic, gram-positive bacterium, with a growth range between 45° C. and 75° C, with an optimum 65°C. Aerobic, rod shaped, endospore forming. 48-58% G+C DNA base composition.	Alliota <i>et al.</i> , 1998, Nazina <i>et al.</i> ,2001
<i>Geobacillus kaustophilus</i> HTA426 ATCC 8005, DSM 7263 ( <i>G.ka</i> )	Mariana Trench, deep-sea sediment.	Thermophilic, gram-positive bacterium, with a growth range between 45° C and 75° C, with an optimum of 60°C. Aerobic, rod shaped, endospore forming.	Takami <i>et al.</i> , 1997
<i>Geobacillus furnas</i> ( <i>G.fs</i> )	Azores, Hot spring run off water. Sampled by Morant (2010)	Self-sampled isolate. Limited available information. 60°C, pH6.0-7.0	n/a
<i>Anoxybacillus flavithermus</i> DSM 2641 ( <i>A.fl</i> )	New Zealand Hot spring	Gram-positive, facultatively anaerobic, endospore-forming, motile, rod-shaped bacterium, with an optimum growth temperature of 55 °C.	Pikuta <i>et al.</i> , 2000
<i>Kosmotoga olearia</i> ATCC BAA-1733, DSM21960 ( <i>K.ol</i> )	North Sea off Norwegian coast, oil production fluid	Thermophilic, heterotrophic gram-negative bacterium. Cells of strain were non-motile rods with a sheath-like structure, or toga. Grew at 20–80 °C (optimum 65 °C), pH 5.5–8.0 (optimum pH 6.8) The G+C content of the genomic DNA was 42.5%	DiPippo <i>et al.</i> 2009
<i>Carboxydotherrnus ferrireducens</i> JW/AS-Y7, DSM11255 ( <i>C.fe</i> )	Yellowstone National Park Hot Springs (USA)	Gram-positive, thermophilic, strictly anaerobic rod-shaped bacterium. Growth temperature 50-74°C with optimum at 65°C. The G+C content of the genomic DNA was 41%. pH 5.5-7.6 with optimum at pH 6.	Slobodkin <i>et al.</i> , 1997

Organism	Isolation	Growth characteristics	Reference
<b><i>Thermodesulfatator indicus</i></b> JCM 11887, DSM15286 ( <i>T.in</i> )	Deep-sea hydrothermal vent site at the Kairei vent field on the Central Indian Ridge	A thermophilic, marine, anaerobic, chemolithoautotrophic, sulfate-reducing bacterium. Cells are non-spore forming Gram-negative motile rods. The temperature range for growth was 55-80°C, with an optimum at 70°C. The pH range for growth was pH 6.0-6.7, with an optimum at approximately pH 6.25. The G+C content of the genomic DNA was 46.0 %.	Moussard <i>et al.</i> 2004
<b><i>Thermodesulfatator atlanticus</i></b> AT1325, DSM 21156 ( <i>T.at</i> )	Deep-sea hydrothermal vent at the Rainbow site on the Mid-Atlantic Ridge	A strictly anaerobic, thermophilic, sulfate-reducing bacterium. Cells are Gram-negative motile rods (approximately 2.4 x 0.6 microm) with a single polar flagellum. . The temperature range for growth was 55-75°C (optimum, 65-70°C), at pH 5.5-8.0 (optimum pH 6.5-7.5). The G+C content of the genomic DNA was 45.6 %.	Alain <i>et al.</i> 2010
<b><i>Thermodesulfobacterium hydrogenophilum</i></b> JCM 11239, DSM14290 ( <i>T.hy</i> )	Deep-sea hydrothermal vent Gulf of California, Guaymas Basin,	A thermophilic, non-spore-forming, marine, sulfate-reducing bacterium, The gram-negative-staining cells occurred singly or in pairs as small, highly motile rods. The temperature range for growth was 50-80°C with an optimum at 75°C. The pH range for growth at 70°C was 6.3-6.8, with an optimum at 6.5. The G+C content of the genomic DNA was 28 mol%.	Jeanthon <i>et al.</i> 2002
<b><i>Marinithermus hydrothermalis</i></b> JCM 11576, DSM14884 ( <i>M.hy</i> )	Deep-sea hydrothermal vent chimney sample collected from the Suiyo Seamount in the Izu-Bonin Arc, Japan	A novel thermophilic marine bacterium. Cells were rod-shaped, occurring in pairs or filamentous, and stained Gram-negative. Growth was observed between 50.0 and 72.5°C (optimum 67.5°C; 30 min doubling time) and at pH 6.25-7.75 (optimum pH 7.0). It is a strictly aerobic heterotroph. The G+C content of the genomic DNA is 68.6%	Sako <i>et al.</i> 2003
<b><i>Thermosediminibacter oceanic</i></b> ATCC BAA-1034, <b>DSM16646</b> ( <i>T.oc</i> )	Core sample of deep sea sediment in Peru.	An anaerobic thermophile. The temperature range for growth was 52-76°C with an optimum at around 68°C. The pH range for growth was from pH 6.3 to 9.3 with an optimum at pH 7.5 . The G+C of the genomic DNA was 46.3%	Lee <i>et al.</i> 2005
<b><i>Geobacillus caldovelox</i></b> ( <i>G.ca</i> )	GeneSys Biotech Ltd. isolate	Thermophilic Gram-positive bacterium, with a growth range between 45° C and 70° C, with an optimum at 60°C. Aerobic, rod shaped, endospore forming.	n/a
<b><i>Geobacillus</i> sp. melbourne</b> ( <i>Gsp.M</i> )	GeneSys Biotech Ltd. isolate	Thermophilic Gram-positive bacterium, with a growth range between 45° C and 70° C, with an optimum at 60°C. Aerobic, rod shaped, endospore forming.	n/a

**Table 3.2**

An overview of the thermophilic strains used in this study.

Eighteen isolates were collected in the Azores (Table 3.1) but only one isolate, labelled 'No.8', collected from thermal water run-off at 58°C (pH6.0) from the 'Furnas' site, showed the presence of a thermostable organism on LB plates. This strain required incubation for 48h at a reduced temperature of 55°C for growth and was labelled '*Gsp. furnas*' (*G.fs*). Two further samples (No.13 and No.15 from Table 3.1), collected from Azorean soil, showed the presence of a moderately thermophilic organism at 45°C. The remaining Azorean samples were noted to have been collected at locations with either a very low pH value (pH1-2) or from extremely high sampling temperatures (70°C-99°C). *G.fs* was the only Azorean sample taken forward for this project.

Soil samples donated from GeneSys Biotech Ltd. revealed the presence of organisms with a variety of thermostabilities, with only sample 'MeIA' (referred to now as *G.me*) showing bacterial growth at 60°C after 24h. The remaining samples were not taken forward for further study with the more thermo-tolerant *Geobacillus*-like organism expected to harbour a more thermostable DNA polymerase enzyme.

From the culture-collected and donated strains, with those isolated from the screened soil and water samples, 12 thermophilic organisms (Table 3.3) were therefore available to search for novel DNA polymerase enzymes potentially suitable for isothermal amplification.

Sample	Organism	Abbreviation
Soil screening	<i>Geobacillus sp melbourneA</i>	<i>G.me</i>
Water screening	<i>Geobacillus sp furnas</i>	<i>G.fs</i>
Donated	<i>Geobacillus kaustophilus</i>	<i>G.ka</i>
	<i>Geobacillus caldovelox</i>	<i>G.cx</i>
Culture collection	<i>Anoxybacillus flavithermus</i>	<i>A.fl</i>
	<i>Kosmotoga oleania</i>	<i>K.ol</i>
	<i>Carboxydotherrmus ferrireducens</i>	<i>C.fe</i>
	<i>Thermodesulfatator indicus</i>	<i>T.in</i>
	<i>Thermodesulfatator atlanticus</i>	<i>T.at</i>
	<b>Thermosediminibacter oceanic</b>	<i>T.oc</i>
	<i>Thermodesulfobacterium hydrogeniphilum</i>	<i>T.hy</i>
	<i>Marinithermus hydrothermalis</i>	<i>M.hy</i>

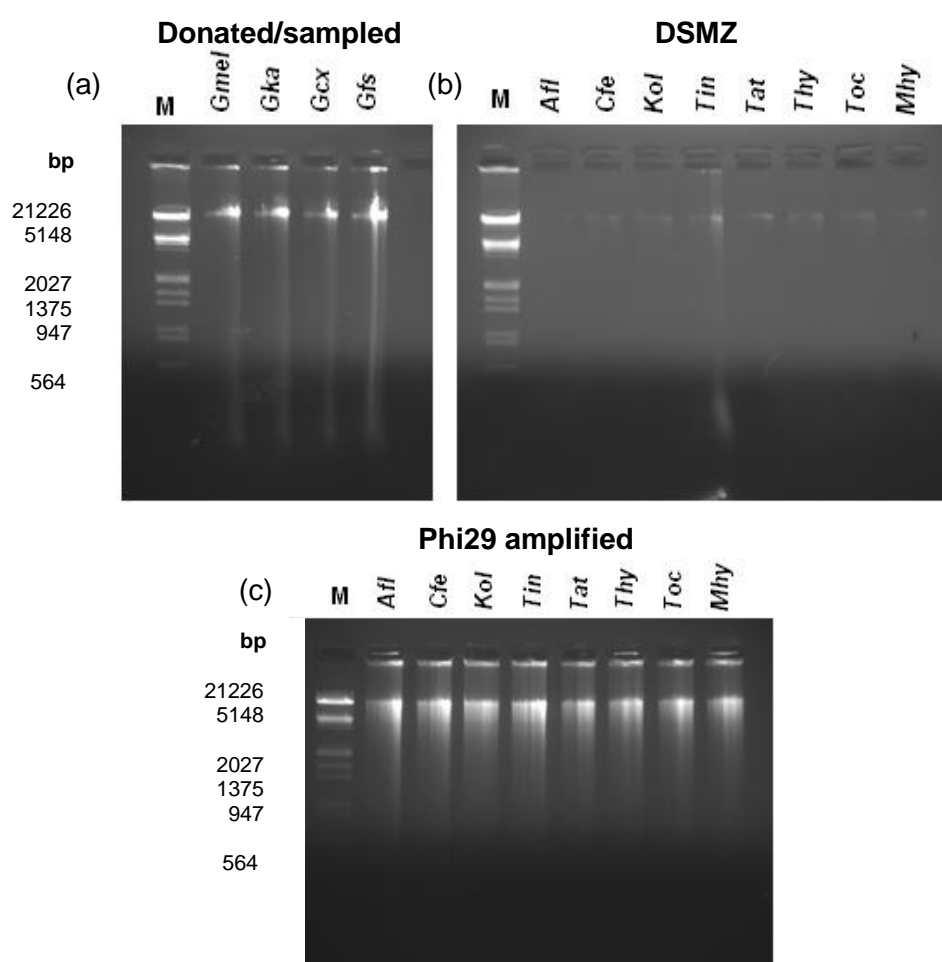
**Table 3.3**

The organisms selected for further study with their abbreviated name used within the thesis text.



### 3.2.2 Isolation of genomic DNA

The genomic DNA (gDNA) was extracted from each of the twelve isolated strains following the method in Chapter 2.2.2. The concentration and purity of gDNA was confirmed by agarose gel electrophoresis (Figure 3.3.a/b). The yield acquired from LB plated (donated and sampled) organisms was shown to be much greater than those from the cultures obtained from the DSMZ culture collection. Where the gDNA concentration was low, the yield was further enhanced using a GenomiPhi™V2 DNA amplification kit (Promega, UK) to generate µg quantities from ng starting concentrations (Fig 3.3c), enabling sufficient DNA to be present for further reactions.

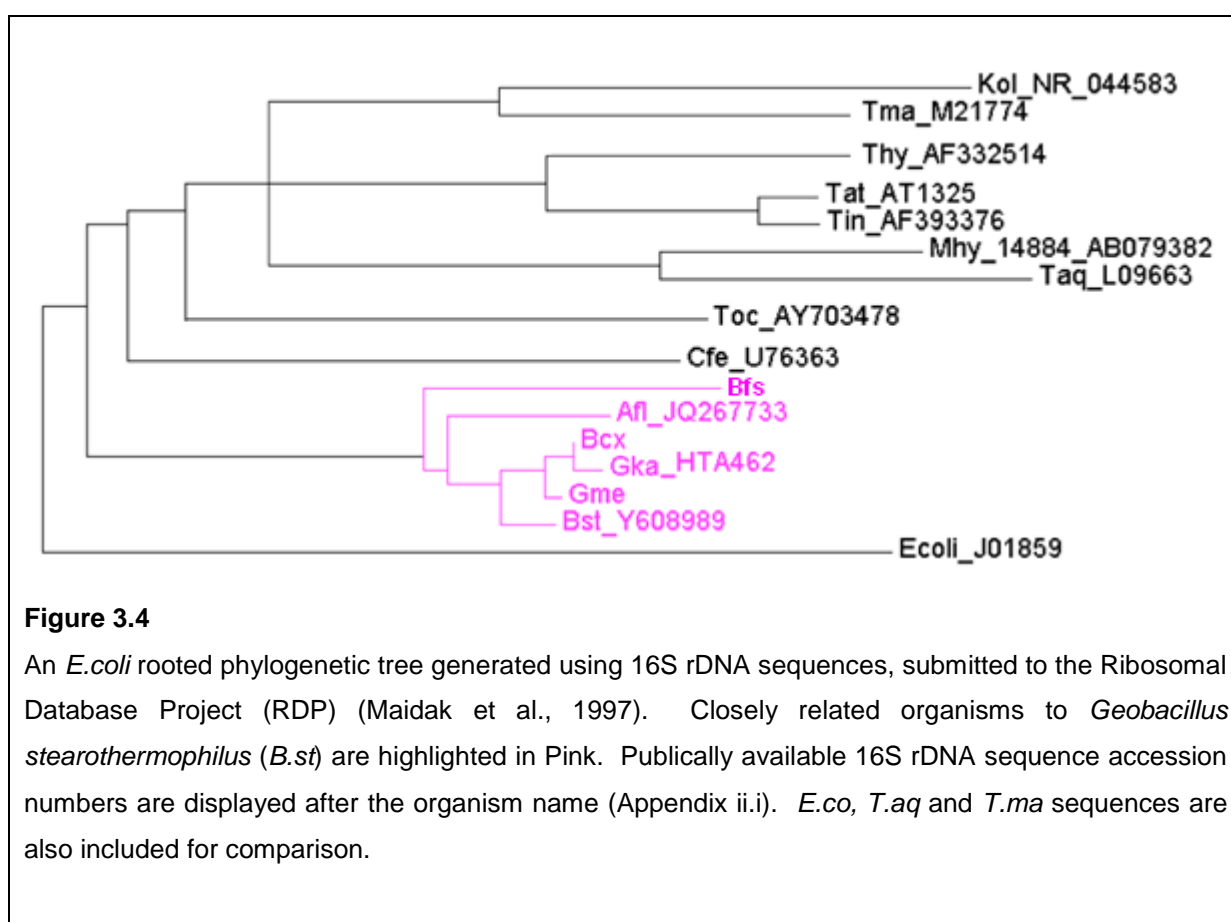


**Figure 3.3**

Genomic DNA isolation visualised on a 1% (w/v) agarose gel with 500ng *EcoR* I/*Hind* III DNA ladder (GeneSys Biotech Ltd. UK) for size comparisons. (a) 1µl gDNA isolated from LB plated strains (100µl total sample). (b) 1µl gDNA isolated from DSMZ strains (100µl total sample). (c) 0.5µl Phi29 amplified gDNA from DSMZ samples (20µl total sample).

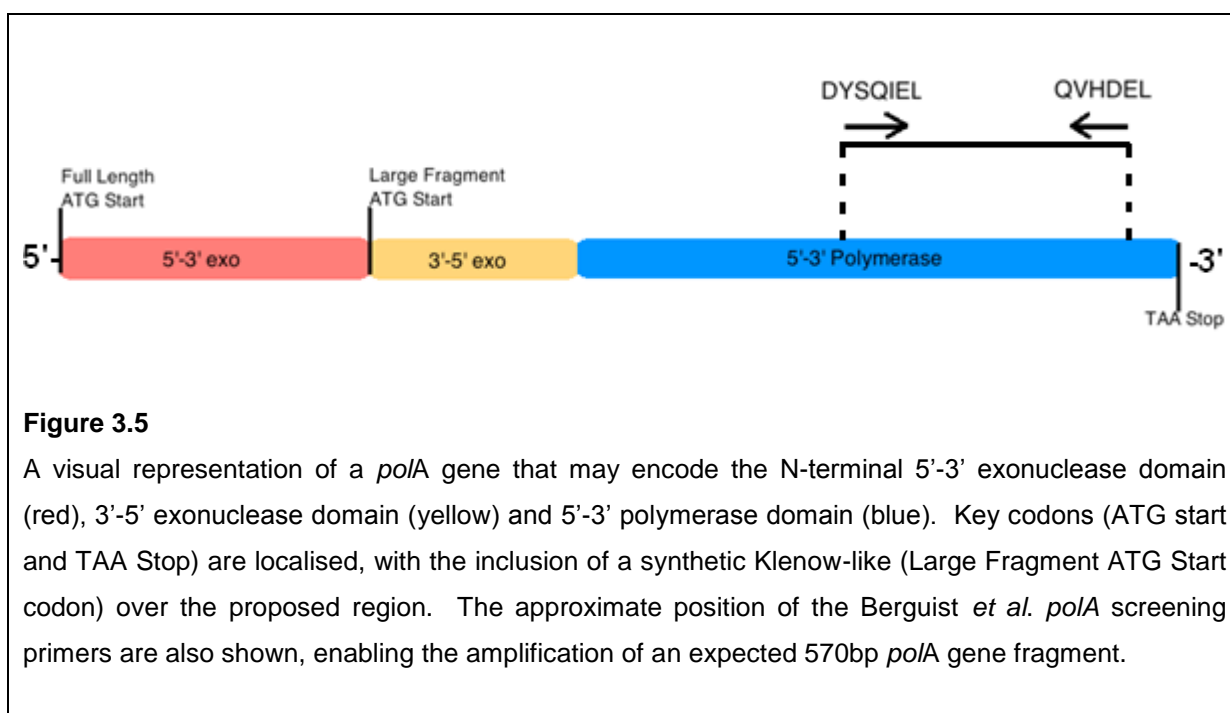
### 3.2.3 16S rRNA gene sequence analysis

To enable taxonomic comparisons of the strains in this study, gDNA was used directly as template in a PCR to amplify the 16S rRNA gene. A single DNA fragment was amplified at approximately 1500bp using the 27f and 1492r universal 16S rRNA primers (Appendix i.i) previously described by Lane *et al.* (1991). The PCR amplicon from each gDNA sample was ExoSAP treated and DNA sequenced using a selection of the 16S\_Universal primers detailed in the Appendix i.i. The DNA sequences (Appendix ii.i) were submitted to the NCBI BLASTN, which confirmed those strains already in the database and identified the newly sampled strains. Strain 'MeIA' was confirmed to be *Geobacillus*, with 99% identity to the 16S rRNA gene of *Geobacillus thermodenitrificans* strain BGSC W9A26. *Gsp. furnas* was shown to share 99% identity to *Bacillus licheniformis* strain HPG16, a non-thermophilic, commonly found, spore-forming mesophilic bacterium, with optimum growth temperature of 30-37°C (Huang, 2012). As such, the *Gsp. furnas* strain was re-classified as a *Bacillus-like organism* and labelled *B. furnas* (*B.fs*) for future reference. The 16S rRNA genes identified were submitted to the Ribosomal Database Project (RDP) for alignment and the generation of a phylogenetic tree (Figure 3.4).



### 3.2.4 *polA* identification

Uemori *et al.* (1993) and Bergquist *et al.* (2004) had reported the successful screening and identification of unknown *polA* genes using degenerate primers based on highly shared *PolA* motif regions. The same technique, comprising the Bergquist *et al.* (2004) *polA* primer sets, was selected to identify the *polA* sequences of the non-sequenced strains in this study. The degenerate forward primers ATF/GCF1/GCF2 (Appendix i.ii) were chosen to screen for the 'DYSQIEL' motif and the reverse primers ATR/GCR were chosen to target the 'QVHDEL' *polA* motif (Figure 3.5).



gDNA was PCR amplified using a combination of the Bergquist *et al.* (2004) designed CODEHOP style primers (Table 3.4), to screen for an expected ~570bp internal fragment of the *polA* gene. Products from the touch-down PCR were visualised using a 2% (w/v) agarose gel. All gDNA samples targeted showed a clean amplified band at approximately 570bp indicating the presence of the *polA* gene. The PCR product was used directly in a ligation reaction to TA clone the fragments into the pCR2.1<sup>®</sup> vector (Life Technologies, UK). The ligated fragments were then transformed into *E.coli* TOP10F' and an aliquot of the transformed cells plated onto LB agar plates. 1mM IPTG and 40µg/ml X-gal were further included in the agar plates for blue-white colony screening of the recombinant colonies. Five white colonies were selected for each and gridded onto a fresh LB agar plate and screened using the colony-screening PCR method, with vector specific M13 Forward and M13

Reverse primers (Appendix i.iii). PCR products of the expected size were visualised by agarose gel electrophoresis (1.5% (w/v) agarose), ExoSAP treated, and sequenced on the ABI Prism 3100XL genetic analyser using the vector specific primers. DNA sequencing confirmed the presence of *polA* gene fragments from each strain.

	CODEHOP Primers	Fragment size (bp)
<b><i>G.me</i></b>	PolATF/PolATR	569
<b><i>G.cx</i></b>	PolATF/PolATR	569
<b><i>B.fs</i></b>	PolATF/PolATR	569
<b><i>C.fe</i></b>	PolATF/PolATR	569
<b><i>T.at</i></b>	PolATF/PolATR	569
<b><i>T.hy</i></b>	PolATF/PolGCR	569
<b><i>T.oc</i></b>	PolATF/PolGCR/	569

**Table 3.4**

Identifying the CODEHOP primers required to amplify the *polA* specific fragment from each of the non-sequenced organisms.

### 3.2.5 *polA* gene-walking

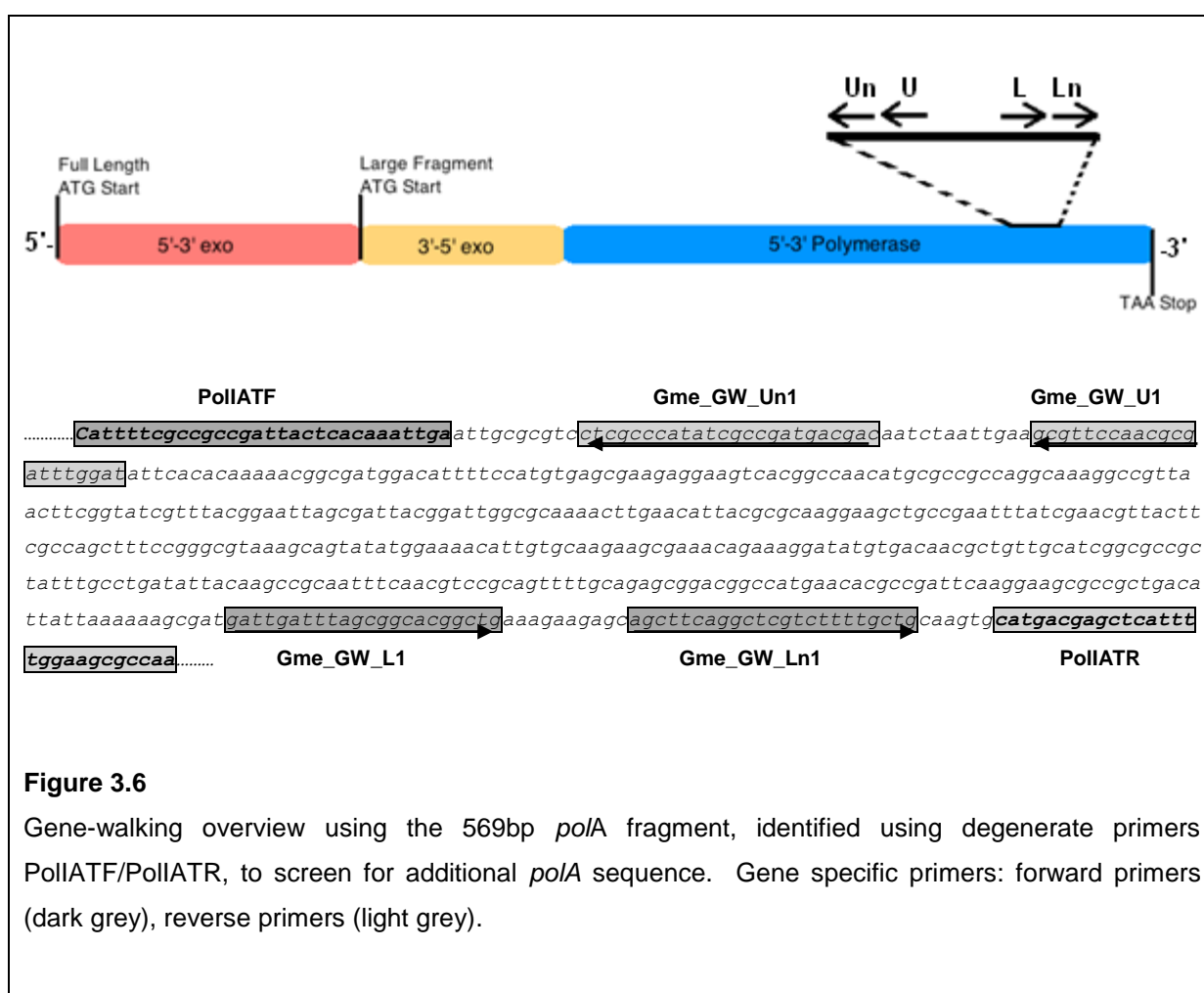
Strain-specific *polA* primers were designed against each sequenced fragment, enabling the gene-walking of the remaining unknown gene sequence. The inverse PCR ‘gene-walking’ method, reviewed earlier (Figure 3.2), relies on two pairs of primers facing out towards the unknown DNA sequence and targeting a circularised template. Primers were designed to ‘walk-along’ a fragmented DNA library of gDNA to reach the 5’ Large Fragment (LF) predicted ATG start and 3’ TAA stop of the DNA *polA* gene. 150ng of gDNA was digested with 10 units of a single restriction enzyme to produce a fragmented DNA library. Twelve individual digest reactions were set up, using a unique 6-cutter restriction enzyme (RE) for each. Restriction enzymes were selected from *Nco* I, *Hind* III, *Eco*R I, *Xba* I, *Bam*H I, *Sal* I, *Pst* I, *Sac* I, *Kpn* I, *Aat* II, *Eco*R V, and *Nde* I to create the library. The use of 6-cutters restricts the frequency of gDNA digestions, potentially allowing larger fragments to be self-ligated. 75ng digested template was self-ligated in a reaction using T4 DNA Ligase to promote circularisation. This reaction provided a fragmented DNA library to be used as template for inverse PCRs for the gene-walking method.

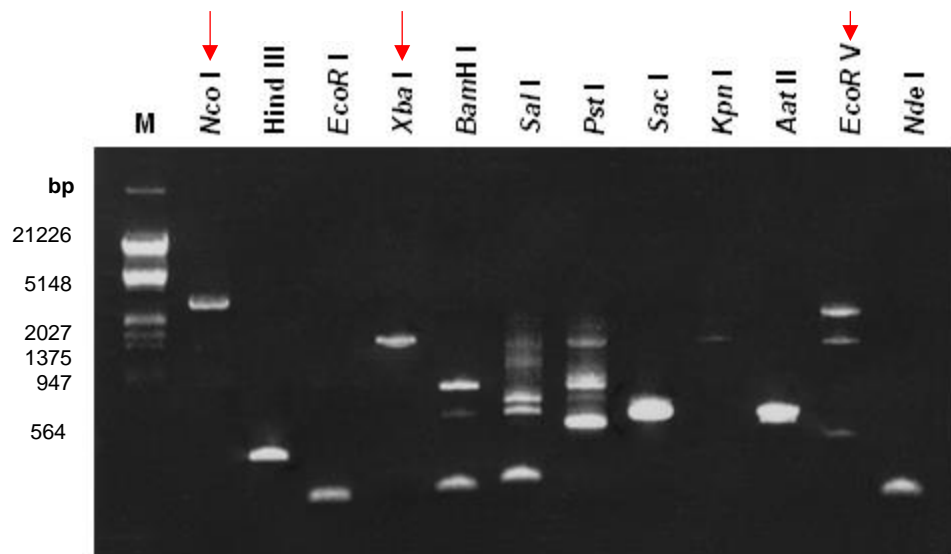
Two rounds of PCR were required for each gene-walking step. Amplified PCR fragments using the first primer pair were diluted 100-fold and used as template in a second (nested) round of PCR. This additional PCR, using nested-specific primers, increased the specificity of the gene-walking method and helped to reduce non-specific products that might be amplified from the gDNA library. Amplified fragments of >500bp were ExoSAP treated and the DNA sequenced using the specific nested primers. DNA sequencing revealed further *polA* sequence data, allowing new gene-walking primers to be designed if required to walk further along the gene.

For simplicity and to save repetition, the identification of *G.me* and *T.at* DNA *polA* are detailed separately below, with the identification of the remaining strains summarised at the end of this chapter.

### 3.2.6 *G.me* LF DNA polymerase I

A BLASTN alignment of the 569bp *G.me* *polA* fragment showed 99% identity to *Geobacillus stearothermophilus* (GenBank: U33536.1), which confirmed the 16S rRNA analysis data as a *Geobacillus*-like organism. Specific gene-walking primers were designed using the *polA* DNA fragment (Figure 3.6) and gene-walking was carried out following the methods defined in Chapter 2.1.11. The nested PCR gene-walking step identified several distinct bands amplified from the majority of digested template libraries (Figure 3.7). Amplicons from the *Nco* I, *Xba* I and *Eco*R V digested template libraries identified the largest fragments with minimal non-specific amplification and these were selected to be ExoSAP treated and DNA sequenced using the nested primers (*Gme\_GW\_UN1* and *Gme\_GW\_LN1*, see Appendix i.iv). The sequence data were subjected to BLASTN analysis and revealed additional *polA* gene sequence, confirming the gene-walking method had been successful.





**Figure 3.7**

*polA* gene-walking using the *G.me* gDNA digest template library. 5µl of PCR product were visualised on a 1.5% (w/v) agarose gel. Red arrows highlight the fragments taken further for DNA sequencing. *E.coR I/Hind III* DNA ladder (M).

Alignments with the published *B.st* Pol I amino acid sequence identified the DNA fragment, derived from the *Xba I* template library, to contain the 3'-terminal stop codon 'TGA' (Figure 3.8). The *polA* sequence derived from the *EcoR V* library revealed 714bp of *polA* sequence towards the N-terminus of the encoded protein, and therefore a further round of gene-walking was required to reveal the *polA* gene sequence to the Large Fragment ATG start position (Figure 3.9).

```

CatTTTcgcgcgcgattactcacaattgaattgcgcgtcctcgccatatcgccgatgacgacaatctaattgaagcgttccaacgcgatttg
gatattcacacaaaaacggcgatggacattttccatgtgagcgaagaggaaagtcacggccaacatgcgccgccaggcaaaaggccgttaacttcg
gtatcgtttacggaattagcgattacggattggcgcaaaacttgaaattacgcgcaaggaaagctgccgaatttatcgaacgttacttcgccag
ctttccggcgctaaagcagtatatggaaaacattgtgcaagaagcgaaacagaaaggatatgtgacaacgctggttcacgcgcgcgctatttg
cctgatattacaagccgcaatttcaacgtcggcagttttgcagagcggacggccatgaacacgcsattcaaggaagcgccgctgacattatta
aaaaagcgatgattgatttagcggcaccgctgaaagaagagcagcttcaggctcgtcttttgctgcaagtgcacgacgagctcattttggaag
cgccaaaagaggaaattgagcgattatgtgagcttgttccggaagtgatggagcaggccgttacgctccgcgtgccgctgaaagtcgactacca
ttacggcccaacatggtatgatgccaaatga
Gme_GW_L1
Gme_GW_Ln1

```

**Figure 3.8**

Gene-walking identified 119bp of *polA* gene sequence (bold italic) including the TGA stop codon using the *Xba I* digested template library.

cgctgcggtggcgaaaatgaaacaatatgaagcgggtgcggtcggatgaagcgggtctatggcaaaggtgccaagcggtcgcttccggacgagcag  
 acgcttgctgagcatctcgttccgaaagcggcagccatttgggcgcttgagcagcggtttatggacgatttgcggaataacgaacaagatcaat  
 tattaacgaagcttgagcagccgctggcagcgatcttggctgaaatggaattcactgggtgaaacgtggatacaaagcggcttgaaacagatggg  
 ttcggagctcgccgaacaactgcgtgccatcgagcagcgcatctacgagctagccggccaagagttcaacattaactcgccaaaacagctcgga  
 gtcatttttatttgaagctgcagctaccggtgctgaagaagacgaaaacaggctattcgacttcggctgatgtgcttgagaagcttgccgcgc  
 atcatgaaatcgctcgaaaacattttgcattaccgccagcttggcaaaactgcaatcaacgtatattgaaggattgttgaaagtgtgcccctga  
 taccggcaagtgcatatcgatgttcaaccaagcgtgacgcaaaactgggcggctcagctcggccgagccgaacttgcaaaacattccgattcgg  
 ctccaagaggggcggaataccgccaagcgttcgtcccgtcagagccggactgggtCattttcgccgcccattactcacaattgaattgcgcg  
 tctcgcgccatatacgccgatgacgacaaatctaattgaaagcgttccaacgcgatttggatattcacacaaaaacggcgatggacattttccatg  
 tgagcgaagaggaagtcacggccaacatgcgcgccaggcgaagggcggttaacttcggtatcggttacggaattagcgattacggattggcgca  
 aaacttgaaacattacgcgcaaggaagctgccgaatttatcgacgttacttcgccagctttccggcgtaagcagtatatggaaaacattgtg  
 caagaagcgaaacagaaaggatatgtgacaacgctgttgcacgcgcgctatttgcctgatattacaagccgcaatttcaacgtccgcagtt  
 ttgcagagcggacggccatgaacacgcgcgattcaaggaagcgccgctgacattattaaaaagcgatgattgatttagcggcacggctgaaaga  
 agagcagcttcaggtcgtcttttgcgtgcaagtgcacgagctcattttggaagcgccaaaagaggaaattgagcgattatgtgagcttgtt  
 ccggaagtgtatggagcagcgcttacgtccgcgtgccgctgaaagtcgactaccattacggcccaacatggtatgatgccaaaTaa

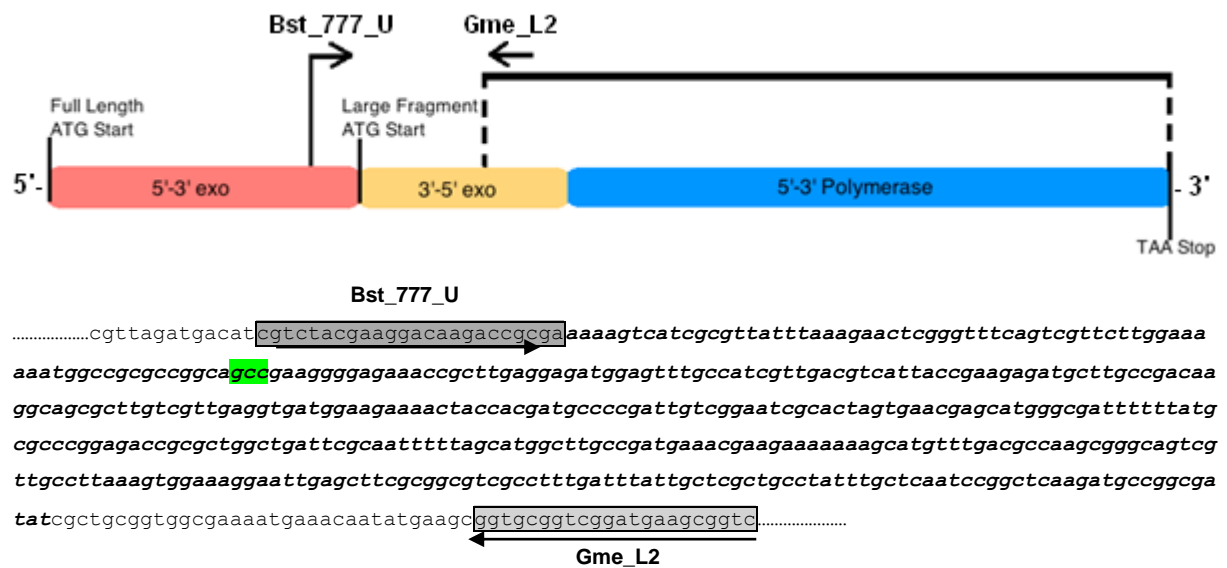
Gme\_GW\_Un1                      Gme\_GW\_U1

**Figure 3.9**

Gene-walking identified 714bp of *polA* gene sequence (bold italic) towards the encoded 5'-terminus using the *EcoR* V digested gDNA template library.

1402bp *polA* sequence, obtained through one round of gene-walking, was submitted to BLASTN and shown to share with 99% identity to *B.st polA* (NCBI\_BSU33536). Primer Bst\_777\_U (Appendix i.iv) was designed 200bp upstream of the LF start position of *B.st* DNA Pol I, to screen for a 508bp fragment by PCR, using the gene specific Gme\_GW\_L2 primer (Appendix i.iv), using the 12 digested libraries as template (Figure 3.10). A 508bp fragment was successfully amplified with these primers from the *BamH* I derived template library, ExoSAP treated, and DNA sequenced using each of the specific primers. BLASTP (Altschul *et al.*, 1990) alignments revealed the proposed LF start position of *G.me* Pol I, when compared to the *B.st* DNA Pol I sequence.





**Figure 3.10**

The PCR screen using Bst\_777\_U and Gme\_L2 identified 508bp of *polA* gene sequence (bold italic) towards the encoded N-terminus using the *Bam*H I digested gDNA template library. The proposed *B.st* DNA Pol I LF start position is highlighted in green.

The 1764bp *G.me* LF *polA* identified shared 99% identity to *B.st polA* (NCBI\_BSU33536). Of the 9bp variation the translated amino acid sequence was shown to differ only by a single residue, at position 433 (Full Length *B.st polA* numbering), from a valine to alanine (Figure 3.11).

```

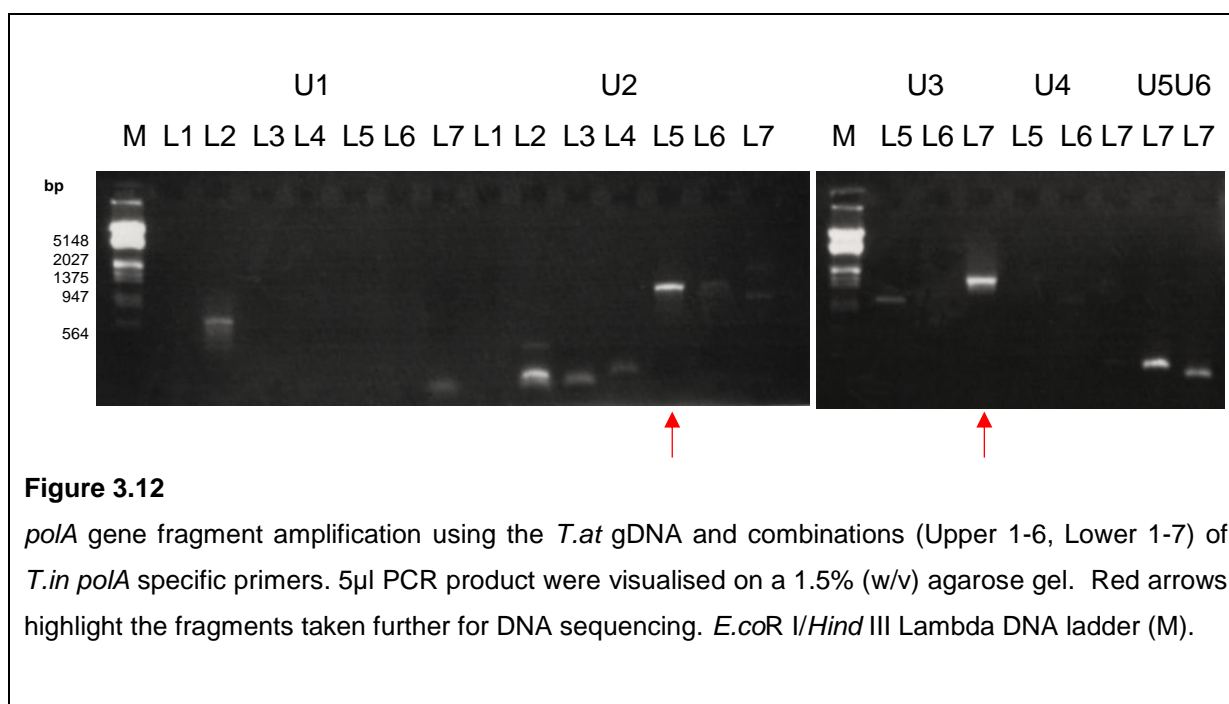
G.me      IAAVAKMKQYEAVRSDEAVYGKGAKRSLPDEQTLAEHLVRKAAAIWALEQPFMDLRLNNE 180
B.st      IAAVAKMKQYEAVRSDEAVYGKGVKRSLPDEQTLAEHLVRKAAAIWALEQPFMDLRLNNE 180
          *****
  
```

**Figure 3.11**

ClustalW multiple sequence alignment (Thompson *et al.*, 1994) of *B.st* LF *polA* (NCBI\_BSU33536) and *G.me* LF *polA*. A single residue difference between the two sequences is highlighted yellow.

### 3.2.7 *T.at* LF DNA polymerase I

The 569bp *T.at* specific *polA* fragment was submitted to BLASTN but no significantly similar sequence was identified in the public database. The DNA sequence was aligned with *T.in* DNA pol I, previously identified by Clark *et al.*, (2009), the only other *Thermodesulfatator*-like organism isolated. Sharing 81% identity, *T.in* *polA* specific primers were predicted to be suitable for the amplification of the remaining *T.at* *polA* gene sequence. *T.in* *polA* specific gene-walking primers disclosed in patent application WO 2009/106795 A1 (Clark *et al.* 2009), were used to target 10ng *T.at* gDNA in a high fidelity PCR. Specific amplicons could be observed using the U2/L5 primer pairs and U3/L7 primers pairs (Figure 3.12). The fragments were ExoSAP treated and DNA sequenced using the same primers used for their amplification.



BLASTN alignments identified *T.at* sequence 168bp from the predicted N-terminal LF start position (based on *T.in* Pol I alignments). The DNA sequencing showed 75% identity to *T.in* DNA *polA* up to the 3'-terminal Stop codon. Additional *T.in* *polA* specific N-terminal primers were designed to be used with C-terminal lower (reverse) primers but no amplification was observed. This suggested that the N-terminal sequence of *T.at* *polA* varies significantly from the *T.in* *polA* DNA sequence, preventing primer annealing. Gene-walking was therefore required to obtain the remaining DNA sequence to reach the predicted *T.at* LF start position.

*T.at* specific gene-walking primers were designed using the 1680bp *polA* sequence so far identified. Primers were designed towards the N-terminal end of the fragment to bias the gene-walking in the direction of the LF start position. No specific amplicons were identified from the PCRs, indicating the gene-walking had been unsuccessful using the 12 digested template libraries described earlier. A further 12 digested template libraries were created, using the restriction enzymes *Nhe* I, *Apa* I, *Bsp* I, *Fsp* I, *Hinc* II, *Cla* I, *Nru* I, *Psi* I, *Pml* I, *Pvu* I, *Spe* I and *Xho* I. Gene-walking was repeated using the *T.at* specific gene-walking primers but, again, no distinct amplified bands were visible by agarose gel electrophoresis.

The gene-walking primers were run in a PCR against themselves, in the presence of *T.at* gDNA, to confirm the correct primer sequence was present. Clean, specific amplification products were amplified. This suggested the primer sequences were correct and the gene-walking method itself was not providing template for amplification. It was suggested the *T.at* gDNA did not contain relevant 6-cutter RE sites close to the desired sequence, preventing circular DNA template libraries.

An alternative method was therefore required to reach the predicted LF start position. Sequence data from a previous work at GeneSys Biotech Ltd. provided 451bp of sequence data upstream of the Full-Length *polA* gene from *Thermodesulfatator indicus* (*T.in*) (unpublished data). This previous work by Clark *et al.* at GeneSys Biotech Ltd. identified the *T.in polA* gene through a similar gene-walking protocol. Gene walking provided sequence data upstream of the Full-Length *T.in* DNA *polA* start position, into the flanking region of the *polA* open reading frame. The 451bp sequence was available to BLASTN but no similarity to the sequence was found in the database. Eight primers were designed randomly against the upstream 451bp DNA sequence and were used in a high fidelity PCR with the *T.in* C-ter STOP primer, previously shown to anneal sufficiently to the *T.at* sequence. Primer Tin(pos-154)U (Appendix i.iv) successfully amplified a 2866bp DNA fragment. The fragment was DNA sequenced, revealing the Full Length ATG start position, and predicted LF start position, based on *KlenTaq* DNA Pol I alignments (detailed later in the chapter).

### 3.2.8 Additional LF DNA polymerase I

An overview of the identification of the *polA* genes from the remaining strains is given below. *polA* genes were gene-walked as previously described, differing with the specific primers (appendix i.iv) and digested gDNA template library used for each strain.

#### ***Geobacillus caldovelox* (G.cx) Pol I**

With the knowledge that the *G.me* DNA *polA* sequence shares 99% identity with the *B.st* *polA* (NEB, UK) sequence, the gDNA of *G.cx* was screened with *Geobacillus* primers, targeting a region upstream of the predicted LF start position, in combination with a *B.st* *polA*-like C-terminal primer (Gcx\_Cter\_L), in an attempt to amplify the *polA* gene. *Geobacillus* DNA sequences were aligned (*B.st*, *G.me*, and *G.ka* *polA*) and a consensus primer was designed labelled 'Gcx\_487\_U' (see Appendix i.iv). It was reasoned that if the *G.cx* *polA* sequence contained the identical primer sequence to those in the alignments, then the primers would successfully amplify the *polA* gene. Primers were designed 487bp upstream of the LF start position.

A high fidelity PCR reaction successfully amplified a 2276bp *polA* sequence. Alignments showed the Gcx LF Pol I to share 99% identity with published *G.ka* LF Pol I sequence.

#### ***Carboxydotherrnus ferrireducens* (C.fe) Pol I**

The internal 569bp fragment obtained through the touch-down PCR showed high identity to *Carboxydotherrnus hydrogenoformans* that had been previously genome sequenced. It was proposed the LF ATG start and C-terminal STOP primers designed from the published sequence may serve to amplify the *C.fe* *polA* sequence, assuming they shared the same sequence over these primer regions. Primers CfeLF\_U and CfeSTOP\_L were designed (Appendix i.v) and run in a high fidelity PCR. The PCR successfully amplified a 1659bp fragment, including the predicted LF start position from the *C.fe* gDNA template. ExoSAP sequencing confirmed the sequence to share 99% identify to *Carboxydotherrnus hydrogenoformans* Z-2901 (NCBI\_CP000141).

#### ***Thermodesulfobacterium hydrogeniphilum* (T.hy) Pol I**

Two rounds of gene-walking, using the primers identified in Appendix i.iv, identified the 2529bp Full Length *polA* sequence.

### **Thermosediminibacter oceani (*T.oc*) Pol I**

Initially the *polA* fragment was successfully gene walked 400bp towards the potential LF start position. During the course of this study, the *T.oc* genome was fully sequenced and became publically available. The internal 1250bp so far identified matched the published *polA* sequence with 100% identity. The published Pol I sequence was aligned with *KlenTaq* Pol I to define the LF start position.

### **3.2.9 Sequenced strains *polA* identification**

A number of strains chosen for this study had been genome sequenced prior to the start of this project, enabling the *polA* gene sequences to be publically available through the NCBI database. The complete *polA* gene sequences were identified for each strain and aligned with the existing *KlenTaq* Pol I and *B.st* Pol I sequences to identify the LF start position.

*polA* sequences were identified for the following DSMZ strains; with detailed gene information given in appendix ii.ii:

#### ***Geobacillus kaustophilus* (*G.ka*) Pol I**

3,544,776 bp genome (BA000043)

#### ***Anoxybacillus flavithermus* WK1 (*A.fl*) Pol I**

2,846,746 bp genome (NC\_011567)

#### ***Kosmotoga oleania* TBF 19.5.1 (*K.ol*) Pol I**

2,302,126 bp genome (CP001634)

#### ***Marinithermus hydrothermalis* (*M.hy*) Pol I**

2,269,167 bp genome (CP002630)

#### ***Geobacillus furnas* (*G.fs*) Pol I**

Re-classified as *Bacillus licheniformis* strain HPG16 (CP000002)

#### ***Thermodesulfatator indicus* (*T.in*) Pol I**

2,322,224 bp genome (CP002683)

Patent: EP2247607B1, US20110008848, WO2009106795A1 (Clark et al., 2009).

### 3.2.10 *PolA* sequence alignments

The *polA* sequences successfully identified using the gene-walking method were aligned against those published and publically available through the NCBI database. Additional Pol I sequences were used in the alignment, including *Thermus aquaticus* (*T.aq*) Pol I and *Thermotoga maritima* (*T.ma*) Pol I, two classical PCR enzymes used widely in biotechnology. Key features of the Pol I sequences are reported in Table 3.5 below. The sequence alignment tool MATGAT (Campanella *et al.*, 2003) was used to compare sequence identities and similarities (Table 3.6). The ClustalW alignment tool was used to compare the Pol I sequences enabling common regions and motifs to be assigned (Figures 3.13 and 3.14).

Organism	LF <i>polA</i> (bp)	LF <i>polA</i> (% GC)	LF Pol I (amino acids)	Isoelectric point (pI)
<i>G.me</i>	1764	51.13	587	5.25
<i>T.in</i>	1851	46.03	615	6.77
<i>G.ka</i>	1764	53.51	587	5.58
<i>A.fl</i>	1764	40.02	587	6.04
<i>K.ol</i>	1851	44.52	616	5.73
<i>M.hy</i>	1659	71.44	552	6.70
<i>B.fs</i>	1764	48.42	587	5.37
<i>G.cx</i>	1764	53.74	587	5.64
<i>C.fe</i>	1659	42.56	552	5.71
<i>T.at</i>	1848	45.35	615	6.77
<i>T.hy</i>	1704	31.16	567	8.73
<i>T.oc</i>	1764	49.55	587	5.88

**Table 3.5**

A summary of the LF Pol I's identified. The GC content and theoretical Isoelectric points (pI) were estimated from the reported LF Pol I sequence using the Molecular Biocomputing Suite (MBCS) application (Muller *et al.* 2001).

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. Bst		99.8	88.6	87.9	69.3	76.7	38.5	46.9	45.9	46.2	39.7	45.1	43.1	42.7	41.6
2. Gme	99.8		88.8	88.1	69.5	76.8	38.5	46.9	45.9	46.2	40.1	45.1	43.1	42.9	41.6
3. Gka	94.0	94.2		98.6	68.5	76.1	39.1	46.6	46.5	46.9	39.7	45.2	42.7	42.6	41.2
4. Gcx	93.9	94.0	99.3		68.0	75.1	38.6	46.3	46.2	46.6	39.7	44.6	42.4	42.3	40.9
5. Bfs	83.3	83.5	82.8	82.5		69.0	39.6	45.1	44.6	45.0	39.5	44.1	41.6	40.1	41.8
6. Afl	85.7	85.9	84.3	84.2	83.3		38.2	46.0	45.1	45.7	39.3	45.6	41.9	40.3	40.6
7. Kol	59.1	58.9	58.8	58.4	57.8	58.4		41.6	45.6	44.4	38.9	40.4	38.8	37.2	59.5
8. Cfe	65.9	65.8	65.6	65.2	65.1	65.6	60.2		44.5	43.6	42.9	46.8	41.9	43.2	44.6
9. Tin	64.1	64.1	64.7	64.6	62.9	62.8	66.6	64.1		86.0	45.7	46.1	42.1	43.9	50.5
10. Tat	64.6	64.6	65.9	65.9	62.9	63.6	66.1	62.8	93.0		45.9	44.3	43.3	42.9	49.2
11. Thy	62.2	62.9	60.8	60.6	62.7	62.0	61.0	64.6	66.3	65.4		41.4	39.9	41.1	40.8
12. Toc	65.8	65.8	66.3	66.1	65.9	66.3	61.7	64.9	64.9	63.4	64.7		42.1	42.9	43.1
13. Mhy	60.8	60.8	59.5	59.5	60.6	58.8	55.0	61.2	60.3	60.0	59.2	60.0		68.6	39.7
14. Taq	60.5	60.6	61.7	61.7	60.8	61.3	56.2	62.7	61.3	60.3	59.8	60.8	81.6		41.0
15. Tma	61.3	61.3	62.0	61.6	60.5	60.2	75.2	62.8	68.9	68.0	61.8	64.3	57.7	59.0	

**Table 3.6**

MATGAT protein sequence alignment comparing the LF Pol I amino acid sequence, using the scoring matrix BLOSUM50. % similarity (yellow), % identity (blue). Also, included in the alignment are *B.st* LF Pol I (NEB), *T.aq* LF Pol I, and *T.ma* LF Pol I.

## ClustalW Pol I amino acid alignment:

	Full length PolA Start	
Tma	-----MARLFLFDGTALAYRAYYALDRSLSTSTGIPTNATYGVARMLVR	44
Kol	-----MANRLFLFDGTAIVYRAYFAIDVSLTNSKGEPTNAVYGTARMLSR	45
Eco	-----MVQIPQNPLILVDGSSSYLYRAYHAFPP-LTNSAGEPTGAMYGVLNMLRS	48
Tin	MAQKSLFPKKLPFKDDKDPIFVIDGSSFVYRAYYAIRGHLSNRKGLPTKAVFGFTQMLLK	60
Tat	MAQKNLFPQKLPFDGKKDPIFVIDGSSFIYRAYFAIKGHLSNRKGLPTKAIFGFTQMLLK	60
Thy	-----MIFLIDGSSFVYRAYFAIPGYLATTKGFPTKAIFGVTQMILK	42
Bst	-----MKKKLVLDGNSVAYRAFFALPL-LHNDKGIHTNAVYGFTMMLNK	44
Gme	-----MKKKLVLDGNSVAYRAFFALPL-LHNDKGIHTNAVYGFTMMLNK	44
Bcx	-----	
Gka	-----MRLKKLVLDGSSVAYRAFFALPL-LHNDKGIHTNAVYGFTMMLNK	46
Afl	-----MQEEVFDLANKLVLDGNSIAYRAFFALPL-LHNDKGIHTNAIYGFTMMLMK	51
Bfs	-----MTERKKLVLDGNSLAYRAFFALPL-LSNEKGIHTNAVYGFTTILMK	46
Toc	-----MKKIMLIDGNSLIHRAFFALPP-LMTSKGVHTNAVYGFMMMLMR	43
Cfe	-----	
Taq	-----MRGMLPLFEPK-GRVLLVDGHHLAYRTFHALKG-LTTSRGEVPQAVYGFAKSLLK	53
Mth	-----MQ-QPSLFDHRPERILIVDGHHLAYRNYFALGE-LTTSRGEVPQAVYGFARTLLK	53
	5'-3' exonuclease	
Tma	FIKDHIIVGKDYVAVAFDCKAATFRHKLLETYKAQRPKTPDLLIQQLPYIKKLVEALGMK	104
Kol	FIKNYIAEG-DYALFAFDRKEATHRHDLFEGYKATRAEMPDALVAQLKYIPDLVEGFGIK	104
Eco	LIMQYKPTH---AAVVFDAKGKTRDELFEHYKSHRPPMPDDLRAQIEPLHAMVKAMGLP	105
Tin	LLREMNPETH---VVVCFDAKGKTRHEMYKEYKANRPPMPDDLVSQIPYIKEVTRAFGVP	117
Tat	LLKEMDPKY---VVVCFDAKGKTRHKVYEEYKANRPPMPDDLAVQIPYIREVTRAFGVP	117
Thy	ILKEWGPEY---IIWFMDEKEPTFRHIAENYKATRPMPEDLKIQIPYIKKIIPALGIP	99
Bst	ILAEEQPTH---LLVAFDAGKTTFRHETFQYKGGQQTPPELSEQFPLRELLKAYRIP	101
Gme	ILAEEQPTH---LLVAFDAGKTTFRHETFQYKGGQQTPPELSEQFPLRELLKAYRIP	101
Bcx	-----	
Gka	ILAEETPTH---MLVAFDAGKTTFRHEAFQYKGGQQTPPELSEQFPLRELLRAYRIP	103
Afl	LIEEEKPTH---ILVAFDAGKTTFRHEVYTEYKGGQKTPPELSEQFPLRELLDAYNIR	108
Bfs	MLEEEKPTH---MLVAFDAGKTTFRHKTfKEYKGGQKTPPELSEQLPFIRELLDAYRIS	103
Toc	ILKEQRPDY---IAVAFDKKSPTRHQEFIEYKANRVRTPEELVGQFDVLKQILKAMNIR	100
Cfe	-MKEIEP-----KIKKEYQEGK-----	16
Taq	ALKEDGDAV---IVVFDAPQPSFRHEAYGGYKAGRAPTPEDFPRQLALIKELVDLLGLA	109
Mth	LLKEDGDCV---IVVFDAPQPSFRHEQFAAYKAQRAPTPEDFKPQLEKIKQLVDLLGLA	109
Tma	VLEVEGYEADDIIAT-LAVKGLPLFDEIFIVTGDKDMLQLVNEKIKVWRIVKGISDLELY	163
Kol	FHSVATLEADDIIAT-AVTKYRDKFDEIVIVSGDKDILQLVDEKVKVLRVSGLDLEEY	163
Eco	LLAVSGVEADDVIGT-LAREAEKAGRPVLISGDKDMAQLVTPNITLINTMTNT----IL	160
Tin	ILEIEGFEADDLIA----AIATRMERPIVIVGGDKDLFPLISEKVVMMWDPMKDE----LI	169
Tat	ILEIEGFEADDLIA----AIATRIDHPVIVGGDKDLFPLISEKVVMMWDPMKDL----FI	169
Thy	VISCPGYEADDLIATFVHKI IKKLDKSAIIVAGDKDLVSLVSEKVCYIDPVREK----LL	155
Bst	AYELDHYEADDIIGT-LAARAEQEGFEVKIISGDRDLTQLASRHVTVDITKKGITDIEPY	160
Gme	AYELDHYEADDIIGT-LAARAEQEGFEVKIISGDRDLTQLASRHVTVDITKKGITDIEPY	160
Bcx	-----	
Gka	AYELENYEADDIIGT-LAARAEQEGFEVKVISGDRDLTQLASPHVTVDITKKGITDIEPY	162
Afl	TYELENYEADDIIGT-LATKAEKEGFDVVIISGDRDLTQLASERIHVHVTTKKGITDMERY	167
Bfs	RYELENYEADDIIGT-LAKSAEKDGFEVKIFSGDKDLTQLATEGTTVAITKKGITDVEYY	162
Toc	YIEIDGYEADDILGS-LSKKAEEAGIFTLIVTGDKDTLQLVSPMVHMLTRKGISEMEIY	159
Cfe	--DLVQFET-----VETEGQIAVVS-----	35
Taq	RLEVPGYEADDVLAS-LAKKAEKEGYEVRIITADKDLVQLLSDRIHVHLHPEGYL-----I	163
Mth	RFELAGYEADDVIGS-LAKKAEAEGYEVRIVTSRDSYQLLSDKVRVLKPDGEE-----V	163



Tma	DAQKVKEKYGVPEPQQIPDLLALTGDEIDNIPGVTGIGEKTAVQLLEKYKDLEDILNHVRE	223
Kol	DRRKVEEKFGLPPEKIYELLSLGGDTSDNIPGVPGIGIKTALKLLKEYGSIDEIYKNIRK	223
Eco	GPEEVVNKYGVPELIIDFLALMGDSSDNIPGVPGVGEKTAQALLQGLGGLDTLYAEPEK	220
Tin	DESWIKKRFGIEPKLLDVRLAGDSIDNVPVPGVGEKTAALRIKEYGSLEEVLNHAEE	229
Tat	DQTWIKERFGVEPEKLLDVRLAGDSIDNIPGVPGVGEKTAALRIKEYGSLEEVLNHAEE	229
Thy	DKEAFFKKYGFDPITFPEFRALTGDKSDNIPGVPGVGEKTAALRIKEYGSLEEVLNHAEE	215
Bst	<b>TPETVREKYGLTPEQIVDLKGLMGDKSDNIPGVPGVGEKTAVKLLKQFGTVENVLASIDE</b>	220
Gme	TPETVHEKYGLTPEQIVDLKGLMGDKSDNIPGVPGVGEKTAVKLLKQFGTVENVLASIDE	220
Bcx	-----	
Gka	TPETVREKYGLTPEQIVDLKGLMGDKSDNIPGVPGVGEKTAVKLLRQFGTVENVLASIDE	222
Afl	TPKHVLEKYGLTPAQIVDLKGLMGDASDNIPGVPGVGEKTAALRIKEYGTVEHVLASLEH	227
Bfs	TPEHVREKYGLTPEQIIDMKGLMGDSSDNIPGVPGVGEKTAIKLLKQFHTVEELLSSIDE	222
Toc	DPDKMAERFGIPPAIPDMKGLMGDSSDNIPGVPGVGEKTAALRIKEYGTVEHVLASLEH	219
Cfe	-----DGIFYDD-----GEKTKFYSLDRLEIQLIEFR----	62
Taq	TPAWLWEKYGLRPDQWADYRALTGDESDNIPGVPGVGEKTAALRIKEYGTVEHVLASLEH	223
Mth	TPETVREKYGVTVAQWVDFRALTGDAASDNIPGVPGVGEKTAALRIKEYGTVEHVLASLEH	223

Tma	LP-----QKVRKALLRDRENAILSKKLAILETNVPIEINW-EELRYQGYDREKLLPLL	275
Kol	LS-----PGLRKKLMGDKKALEMSKKLVRLVTDADLHIEL-EELRYRGLDKEKRLREL	275
Eco	IAGLSFRGAKTMAAKLEQNKEVAYLSYQLATIKTDVELELTC-EQLEVQQPAAEELLGLF	279
Tin	IK-----QKRLRENLIKHAAGDALISKKLVELEAKAPIPLEP-DFYRKRPLNALKLREL	282
Tat	IK-----QKRLRENLIKHAAGDALISKKLVELEAKAPIPLEP-DFYRKRPLNALKLREL	282
Thy	LT-----PLKLRNLLKYKEQVLLNNVALLTLNFNAPLPSESLEYERKDPDYNELKKLF	269
Bst	<b>VK-----GEKLKENLRQHRDLALLSKQLASICRDAPVELSL-DDIVYEGQDREKVIALF</b>	273
Gme	VK-----GEKLKENLRQHRDLALLSKQLASICRDAPVELSL-DDIVYEGQDREKVIALF	273
Bcx	-----	
Gka	IK-----GEKLKETLRQHREMAALLSKKLAIRRDAPVELSL-DDIVYQGEDREKVVALF	275
Afl	IS-----GKKLKENLQTYHEQALLSKQLATIRRDVPLTSL-DELAHWSYDAERVAALF	280
Bfs	VS-----GKKLKEKLEEFKEQALMSKELATITTEAPLEVSL-DSLGYEGFDRVAVKIF	275
Toc	LK-----G-KLRNLIKYEQARVSKRLATIVRDVELDVDL-EEIALTEPDYAEELLKIF	271
Cfe	-----	
Taq	LKP-----AIREKILAHMDDLKLSWDLAKVRTDLPLEVDFAK---RREPDRERLRAFL	273
Mth	VTPP-----SVRKKLEAGREKAALSRLSEIHTDLAIEVDFAAC-HRRPVDREALRAFL	276

	<b>Large Fragment Start</b>	<b>3'-5' exonuclease domain</b>	
Tma	KELEFASIMKELQLY----EESPEVG-----YRIVKDLVEF		307
Kol	LKLEFASLLREFDLTNLGSNNEDNG-----YRPVSSQOEL		311
Eco	KKYEFKRWTADVEAGKWLQAKGAKPAKPQETSVADEAPEVTATVISYDNYVTILDEETL		339
Tin	LELEFKLLKELPATKTLSDYQ-----YELVLDPEKDV		314
Tat	LELEFKLLKELPATKTISYDE-----YELVTDSSRM		314
Thy	KELEFRKFLTELKIPS-----AELKARLVEE		295
<b>Bst</b>	<b>KELGFQSFLEKMAAPAAEAGEK-----PLEEMEFIVDVI</b>		307
Gme	KELGFQSFLEKMAAPAAEAGEK-----PLEEMEFIVDVI		307
Bcx	-----MEEK-----PLAKMAFTLADRV		18
Gka	KELGFQSFLEKMESESEEEK-----PLAKMAFTLADRV		309
Afl	QELGFTSLMDKIGQ-SSQEQL-----SLTDISFVTVQTI		313
Bfs	KDLGFHSLLERIGEEAGEKEEE-----QLEEIDVMIKTDI		310
Toc	RELEFYTLINKLPRPQEEKEHP-----EKLSTVIDYSGF		306
Cfe	-----SFLKEIEPKIKK-----EYQEGKDL		20
<b>Taq</b>	ERLEFGSLLHEFG-----LLESPKAL		294
Mth	EALFEGSILRELG-----LIEA-RSA		296

**Asp355 Glu357**  
**3'-5' exo motif I**

Tma	EKLIKLRSPSFAIDLETSSLDPFDCDIVGISVSFKPKAEYIPLHHRN---AQNLD	364
Kol	DELFETLKGSDVISIDLETSSLDPHEAKIVGISVAIKDGEGFYIPVNHESS--DWQADEK	369
Eco	KAWIAKLEKAPVFAFDDETDSLNDNISANLVGLSFAIEPGVAAYIPVAHDYLDAPDQISRE	399
Tin	KEIVEKAKGAEVVAIDLES DTKDPMRGKIVGVS LCFNPPKAYYFPRHEGL-EAQKQLPW	373
Tat	HEILAKAREKGLVVIDLESNHIDPMRGKIVGVALCFEPPKAYYFPRHEGL-EARKQLPW	373
Thy	KIDLEKLKKAELISIKIQGN-----LFSLTNPEELTVASDEK-EAYK-LSV	339
Bst	TEEMLADKAA--LVVEVMEENYH--DAPIVGIALVNEHG-RFFMRPETALA-----DS	355
Gme	TEEMLADKAA--LVVEVMEENYH--DAPIVGIALVNEHG-RFFMRPETALA-----DS	355
Bcx	TEEMLADKAA--LVVEVMEENYH--DAPIVGIAVNEHG-RFFLRPETALA-----DP	66
Gka	TEEMLADKAA--LVVEVMEENYH--DAPIVGIAVNEHG-RFFLRPETALA-----DP	357
Afl	DEHMLTKEGA--LVVEVL DENYH--QAPIVG FALVNERG-HFFIPTDIALA-----SS	361
Bfs	TDDL FASPAS--LVVEQLGDN YH--EAPILGFSIVNEHG-AFFIPEETAVQ-----SD	358
Toc	GRMMERVRAAGVLAVELKTDGRNPM DAHLIGIGFSPSRGEGFYVPAEVLEK-----SP	359
Cfe	VQVETVETEG--QIA--VVFSDG FVDDGEKTKFYSLDRLNEIQEIFRNKK-----67	
Taq	EEAPWPPPEG--AFVGVFLSRKEPMWADLLALAAARGGRVHRAPEPYKALR-----343	
Mth	EEAPWPPPE--AFLGYVLD RQPMAELKGLAGAWEGRVARGPARAKELA-----345	

	<b>3'-5' exo motif II</b>	<b>3'-5' exo motif III</b>	
Tma	EVLKKLKEILEDPGAKIVGQNLKFDYKVL MVKGVEPVPPYFDTMIAAYLLEPNEKKFNLD		424
Kol	QVLLRLKELLEDSG TKIVGQNLKFDYEILEKHGIEPVVPHFDTMIAAYLLNPDSRRFNLD		429
Eco	RALELLKPLEDEKALKVGQNLK YDRGILANYGIELRGIAFDTMLESYILNSVAGRHDMD		459
Tin	EAFTHLASLIEDPSVKKIGHN IKYDLIILARYGVTLKGLEGDTMLASYLLDPTRRTHGLD		433
Tat	EAFGDIAALIEDQVKKIGHN IKYDLILLARYGVALKGLEGDTLLASYLLNPTRRTHGLD		433
Thy	EVFGN--LKKNAKSEFILHDHKS---FLKNYEIFLD-FLFDTKLAAYLLNP SLKNYDLD		392
Bst	----QFLAWLADETKKKS MFD AKRAVVALKWKGIELRGVAFD LLLAAYLLNPAQDAGDIA		411
Gme	----QFLAWLADETKKKS MFD AKRAVVALKWKGIELRGVAFD LLLAAYLLNPAQDAGDIA		411
Bcx	----QFVAWLGD ETKKKS MFD SKRAAVALKWKGIELCGVSFD LLLAAYLLDPAQGVDDVA		122
Gka	----QFVAWLGD ETKKKS MFD SKRAAVALKWKGIELCGVSFD LLLAAYLLDPAQGVDDVA		413
Afl	----RFKRWLEDEQCKKS VFD AKRAI VALKWN GIELKGVDFD LLLAAYLLNPTDANGDVA		417
Bfs	----CFKEWAEDESKKK WVFD AKRAAVALRWRGIELKGA EFDVLLAAYI INPGHSYDDVA		414
Toc	EVKSDLKAVLADPGITKII HDGKYARTVLAKI GMDFV-YNFD TMLAAYLLDPSKPRYDLE		418
Cfe	-----IITDD----AKGIYHVCLEKGLTFPEVCFDARIAAYVLN PADQNPG LK		111
Taq	-----DLKEARGLLAKDLSVLALREGI GLPPGD-DPMLLAYLLDPSNTTP--E		388
Mth	-----RFEAVHALQAKDLTVWARREGVRVQPGE-DPLLAYLYDPTNSDP--A		390

: \*: :

Tma	DLALKFLGYKMTSYQELMSFSFPLFGFSFADVPVEKAANYSCEDADITYRLYKTL SLKLH	484
Kol	DLALKFLGYRTTSFS SELMNKNQ--LKDQFEKVSVEEAAKYSVEDADIALRLYRVL SKKLY	487
Eco	SLAERWLKHKTITFEEIAGKGK--NQLTFNQIALEEAGRYAAEDADVTLQLHLKMWPD LQ	517
Tin	ELAEVVLGHTMISYKEVTKE LAK--GESFARVPLEKAKVYACEDAHVTYLLYQYFWPKLK	491
Tat	ELAEIILGHCMISYKEVTKE LAK--GESFARVPLEKAEDYACEDAHVTYLLYQYFWPKLK	491
Thy	FLLQEHLDINLNSLKASKNELQA--IKTCG-----LYLLGKDLKKKVE	433
Bst	AVAKMKQYEAVRSDEAVY GKGAK----RSLPDEQTLAEHLVRKAAAIWALEQPFMDDLR	466
Gme	AVAKMKQYEAVRSDEAVY GKGAK----RSLPDEQTLAEHLVRKAAAIWALEQPFMDDLR	466
Bcx	AAAKMKQYEAA RPDEAVY GKGAK----RAVPDEPVLAEHLVRKAAAIWALERPFLDEL R	177
Gka	AAAKMKQYEAVRPDEAVY GKGAK----RAVPDEPVLAEHLVRKAAAIWELERPFLDEL R	468
Afl	AVAKTKQYTDVQSDEEVY GKGAK----QAIPPTNVLAEHLVRKAKAIASLKETYIQELK	472
Bfs	SVAKEHQLHIVSADEAVY GKGAK----QAVPDEKELADHLARKAKAISLLREKL LDELE	469
Toc	SVVFDNLGVELKG-----TEDPGRRVAYLIPLKEIMSEK LK	454
Cfe	GLYLKYD-----LPVYEDVSLNIRGLFY LKKEMMKKIF	142
Taq	GVARRYG-----GEWTEEAGERAALSERLFANLWGRLE	421
Mth	ATVRRYGA-----GDWSEDPAARALAAAE LWRILGERLA	424

\* :

**Thumb domain**

Tma	EAD-LENVFKIEMPLVNVLMARMELNGVYVDTEFLKKLSEYEGKKLEELAEIYRIAGEP	543
Kol	ESD-LDNIFHKIEMELIPVLAELNGLVYMNVESLKNLSEYEEKRLGKIREELFELAGEP	546
Eco	KHKGPLNVFENIEMPLVPVLSRIERNGVKIDPKVLHNSSEELTLRLAELEKKAHEIAGEE	577
Tin	EES-LWKVFTEIDRPLIEVLAHMEMVGIKIDTAYLRGLSREMAEKLKELEEKIYTLAGEK	550
Tat	EES-LWRVFEEIERPLIKVLARMEMAGIKIDVPYLRALSQELAQKLKELEQKIYEIAQEQ	550
Thy	EEN-LTTWLKKVEIPLSEVLFEMEKKGFKIDIEYVRELNQRYQKRLKEIEEKLFEIAGFR	492
Bst	<b>NNE-QDQLLTKEQLPLAAILAEMEFTGVNVDTKRLEQMGSELAEQLRAIEQRIYELAGQE</b>	525
Gme	NNE-QDQLLTKEQLPLAAILAEMEFTGVNVDTKRLEQMGSELAEQLRAIEQRIYELAGQE	525
Bcx	RNE-QDRLLVELEQPLSSILAEMEFAGVKVDTKRLEQMGSELAEQLPTVEQRIYELAGQE	236
Gka	RNE-QDRLLVELEQPLSSILAEMEFAGVKVDTKRLEQMGKELAEQLGTVEQRIYELAGQE	527
Afl	RNE-QFELLVHLELPLTFILAQMEFYGVKVDVDRLEQMGKEFTAQLEQIEQRIYELAGTT	531
Bfs	ENE-QLELFEALEMPLAHILGEMESIGVQVDVDRLLKMGEELSAKLAIEYKKIHESAGET	528
Toc	SCA-MEELFFGVEMPLSFVLSDMEMTGKVDPEKLESLSREFGEKLEELTGEIYRLAGVE	513
Cfe	EQE-QERLFYEIELPLTPVLAQMEHTGIQVDREALKEMSLELGEQIEELIREIYALAGEE	201
Taq	GEERLLWLYREVERPLSAVLAHMEATGVRLDVAYLRALSLEVAEEIARLEAEVFRLAGHP	481
Mth	GEEALWWLYREVERPLAGVLAEMEHAGVRVDVAYLEALSDELGREIAAIEAEVHRLAGRA	484

:: \* : \* : \* \* . : : . . . : : . . \*

**Motif 1**

Tma	FNINSPKQVSRIIFEKLGKIPRGKTTKTGDYSTRIEVLEELAGEHEIIPILILEYRKIQKL	603
Kol	FNPNSPTQVSKILFEKLGKLNPP-KKTKHGAYSTAAVLEELVNEHPIIQKLLDYRKYQKL	605
Eco	FNLSTKQLQTILFEKQGIKPL-KKTPGGAPSTSEEVLEELALDYPLPKVILEYRGLAKL	636
Tin	FNINSSKQLGQILFEKLGKLPVKKTPKKTAYSTDNEVLEELSAVHELPRILEYRTLAKL	610
Tat	FNINSSRQLAHILFEKLGKLPVKKTPKKTAFSTDNEVLEELSTLHELPRILEYRTLAKL	610
Thy	FNPRSSREVGIVLFEKLGKLPVKKTPKSSLPSTDAEVLEELAPLHPFVKLLQYRTLYKI	552
Bst	<b>FNINSPKQLGVILFEKLGKLPVKKTKTG--YSTADVLEKLAPHHEIVENILHYRQLGKL</b>	583
Gme	FNINSPKQLGVILFEKLGKLPVKKTKTG--YSTADVLEKLAPHHEIVENILHYRQLGKL	583
Bcx	FNINSSKRLGVILFEKLRPLVKKTKTG--YSTADVLEKLAPYHEIVENILHFRQLGKL	294
Gka	FNINSPKQLGVILFEKLGKLPVKKTKTG--YSTADVLEKLAPYHEIVENILHYRQLGKL	585
Afl	FNINSPKQLGTILFEKLGKLPVKKTKTG--YSTADVLEKLAPYHEIIEQILHYRQLGKL	589
Bfs	FNINSPKQLGVILFDKLGKLPVKKTKTG--YSTADVLEKLDRKHVIEDILHYRQLGKL	586
Toc	FNINSPKQLGEVILFEKLNLPVKKKSG--YSTDAEVLEKLKNAHPVVEKILEYRFLMKM	571
Cfe	FNLNSPKQLGVILFEKLGKLPVKKTKTG--YSTDAEVLEELLPYHEIIGKILNYRQLMKL	259
Taq	FNLNSRDQLERVLFDGLGPAIGKTEKTGKRSTAAVLEALREAHPIVEKILQYRELTKL	541
Mth	FNLNSRDQLEVILYDELGLTPTRRTQKTGRRSTAAALEALVGHPIVERILAYRELSKL	544

\*\* \* . : : \* : : : : : . \*\* . \*\* \* : . : \* : \* \* :

**Palm domain      Motif 2      Arg629**

Tma	KSTYIDALPKMVNPKTGRIHASFNQGTGTATGRLSSSDPNLQNLPTKSEEGKEIRKAIVPQ	663
Kol	KSTYLDTLPTLVNPTGRIHASYHQTGTGTGRLSSSNPNMQNLPIKGEEGKEIRKCVVPQ	665
Eco	KSTYTDKPLPLMINPKTGRVHTSYHQAATATGRLSSSDPNLQNLIPVRNEEGRRIRQAFIAP	696
Tin	KSTYVDALPKMVNPETGRLHTSFNQTVTATGRLSSSDPNLQNLIPVRGEEGLKIRQAFVPE	670
Tat	KSTYVDALPKMANPETGRIHTSFNQTTATGRLSSSDPNLQNLIPVRGEEGTKIRKAFVPE	670
Thy	KSTYLEAFLKYVSSSDWRLHTEFNQGTGTATGRLCSQNPNLQNLIPKGDGLAIRRVFISE	612
Bst	<b>QSTYIEGLLKVVVRPDTGKVHTMFNQALTQTGRLSSAEPNLQNLIPIRLEEGRKIRQAFVPS</b>	643
Gme	QSTYIEGLLKVVVRPDTGKVHTMFNQALTQTGRLSSAEPNLQNLIPIRLEEGRKIRQAFVPS	643
Bcx	QSTYIEGLLKVVVRPDTKKVHTIFNQALTQTGRLSSSTEPNLQNLIPIRLEEGRKIRQAFVPS	354
Gka	QSTYIEGLLKVVVRPDTKKVHTIFNQALTQTGRLSSSTEPNLQNLIPIRLEEGRKIRQAFVPS	645
Afl	QSTYVEGLMKVVRKDTGKVHTIFNQALTQTGRLSSSTEPNLQNLIPIRLEEGRKIRQAFVPS	649
Bfs	QSTYVEGLLKVIKKDSHKVHTIFNQALTQTGRLSSSTEPNLQNLIPIRLEEGRKIRQAFVPS	646
Toc	KSTYADGLLALVDKSTYRIHSNFNQTTATGRISSTEPNLQNLIPVKTDIGRKIRGVFAE	631
Cfe	KSTYTDGLMPLINERTGKLHTTFNQGTTLTGRLASSEPNLQNLIPIRLELGRKLKRMFIPS	319
Taq	KSTYIDPLPDLIHPRTGRLHTRFNQTTATATGRLSSSDPNLQNLIPVRTPLGQIRIRAFIAE	601
Mth	KGTYLDPLPRLVHPATGRIHTRYHQTGTATGRLSSSDPNLQNLIPVTEVGRRIRRAFAE	604

:. \*\* : : : : : \* \* \* : \* : \* : \* : \* : \* : \*

**Asp653** **Arg702**

**Motif A Fingers domain**

Tma	DPNWWIVSADYSQIELRILAHLSGDENLLRAFEEGIDVHTLTASRIFNVKPEEVTTEMR	723
Kol	KSDWKIISADYSQIELRVLAHFSEDERLISAFKNGEDVHALTASRLYGVSVKDVTPEMRQ	725
Eco	-EDYVIVSADYSQIELRIMAHLSRDKGLLTAFAGKDIHRATAAEVFGLPLETVTSEQRR	755
Tin	-EGFLFLSADYSQIDLRVLAHYSGETLIKAFWQGEDIHRRATAAEIFGIPPEEVTPEMR	729
Tat	-KGALFLSADYSQIDLRVLAHYSKDKNLIEAFKRGEDIHRRATAAEIFGVSPPEVTPEMR	729
Thy	-EGYLLCSLDYSQIELRILAHFSEDENLIRAFERGEDIHFTTACEVFGVPLEKVTPEMR	671
Bst	EPDWLIFAADYSQIELRVLAHIADDDNLIEAFQRDLDIHTKTAMDFHVSEEEVTANMR	703
Gme	EPDWLIFAADYSQIELRVLAHIADDDNLIEAFQRDLDIHTKTAMDFHVSEEEVTANMR	703
Bcx	ESDWLIFAADYSQIELRVLAHIAEDDNLMEAFRRDLDIHTKTAMDFQVSEDEVTPNMRR	414
Gka	ESDWLIFAADYSQIELRVLAHIAEDDNLMEAFRRDLDIHTKTAMDFQVSEDEVTPNMRR	705
Afl	SDDWVIFAADYSQIELRVLAHIANDENLIAAFHHDLDIHTKTAMDFHVKEDEVTAHMR	709
Bfs	QKGWLIFAADYSQIELRVLAHISKDKNLIEAFNDMDVHTKTAMDVHVSPEEVTPEMR	706
Toc	SPEHVLLSGDYSQIELRVLAHLSGDEGLIEAFIKGEDIHTRTASEVFGVPPEQVTPLLRD	691
Cfe	PGYDYIVSADYSQIELRLLAHFSEEPKLEAYQKGEDIHRTASEVFGVSLEEVTPEMRA	379
Taq	-EGWLLVALDYSQIELRVLAHLSGDENLIRVFQEGRDIHTETASWMFGVPREAVDPLMR	660
Mth	-PGYVLVAADYSQIELRVLAHLSGDENLKRVFQERRDIHTQTASWMFGVPEAVDPFRR	663

: : \*\*\*\*\*:\*\*\*:\*\*\* : : \* .: . \*: \* \*\* :: : . \* \*

**Lys706** **Tyr714**

**Motif B Ser769 Tyr771**

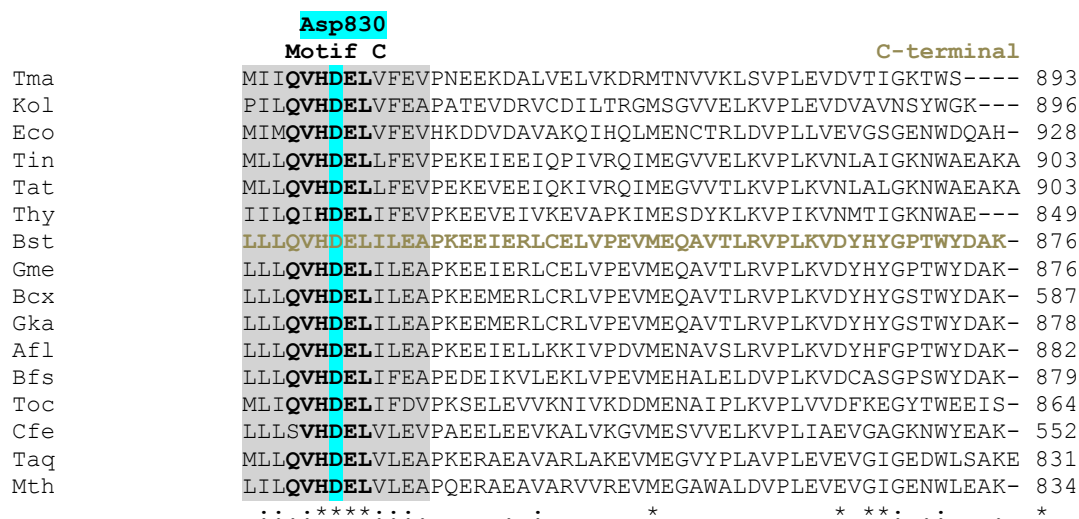
Tma	AGKMVNFSIIYGVTPLYGLSVRLGVPVKEAEKMIVNYFVLYPKVRDYIQRVVSEAKEKGYV	783
Kol	VGKMVNFSIIYIGISPYGLARRLKIKTHIAENMISNYFNAYPGVRKFINEVIHEAKEKGYV	785
Eco	SAKAINFGLIYGMSAFGLARQLNIPRKEAQKYMDLYFERYPGVLEYMERTRAQAKEQGYV	815
Tin	MAKTINFGIVYGMSPYGLAKELKIGRREAKAFIERYFERYPGVKRYMEQIVAEAREKGYV	789
Tat	MAKTINFGIVYGMSPYGLAKELKIGRREAKAFIERYFERYPGVKRYMEQIVAEAREKGYV	789
Thy	MSKAINFGIAYGMSAYGLSKELRISVKEAEAIINRYFSRYPKIKEYIEKTIEFARENGYV	731
Bst	QAKAVNFGIVYGISDYGLAQNLNITRKEAAEFIERYFASFPGVKQYMENIVQEAQKQGYV	763
Gme	QAKAVNFGIVYGISDYGLAQNLNITRKEAAEFIERYFASFPGVKQYMENIVQEAQKQGYV	763
Bcx	QAKAVNFGIVYGISDYGLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAQKQGYV	474
Gka	QAKAVNFGIVYGISDYGLAQNLNISRKEAAEFIERYFESFPGVKRYMENIVQEAQKQGYV	765
Afl	QAKAVNFGIVYGISDYGLSQNLGITRKEAAEFIERYFRSYPGVKRYMEEVVQDAKQKGYV	769
Bfs	QAKAVNFGIVYGISDYGLSQNLGITRKEAAEFIERYFHSFQGVKEYMEETVQEAQKQGYV	766
Toc	RAKAVNFGIIGISDYGLAQNLGISTAEAREYIENYLNRYPKVRDYIRETIRNARMSGYV	751
Cfe	HAKSVNFGIVYGISDFGLGRDLKIPREVAGKYIKNYFANYPKVREYLDLVRTAREKGYV	439
Taq	AAKTINFGVLYGMSAHRLSQELAIPEYEAQAFIERYFQSFPKVRWIEKTLEEGRRRGYV	720
Mth	AAKTINFGVLYGMSPHRLSRELGIYAEAEERFIQRYFESYPRVQAYIERTLEQAREKGYV	723

.\* :\*\*\*: \*\*:: . \* . \* : \* : \* : : : : : . : : \*\*\*

**Arg841** **Motif 6**

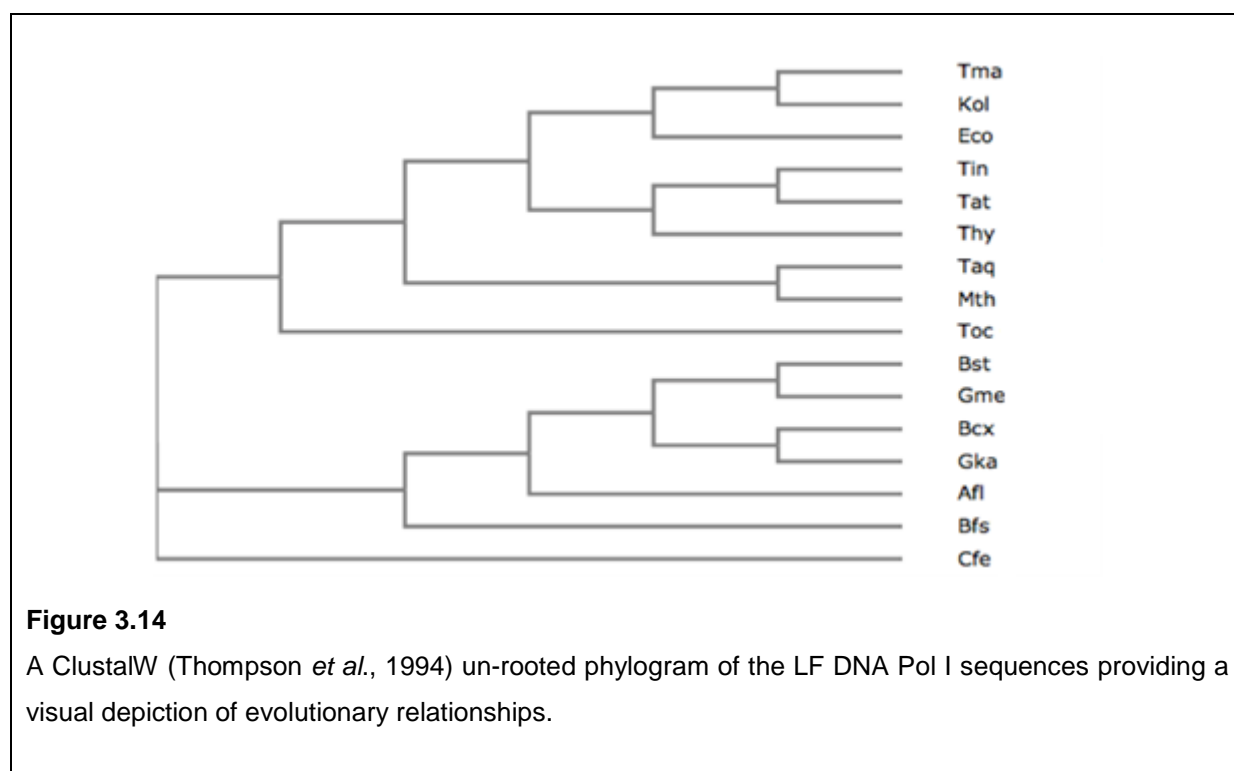
Tma	RTLFGKRKRDPQLMARDNRTQAEGERIAINTPIQGTAAADI IKLAMIEIDRELKERKMRSK	843
Kol	RTLFGKRKRDPHFRTRNKMKIQEGERIAINTPIQGTAAADIMKLAMIKIHKMVKEEGLEAF	845
Eco	ETLDGRRLYLPDIKSSNGARRAAAAERAAINAPMQGTAAADI IKRAMIAVDLWLAQEQPRVR	875
Tin	ETLFGKRKRPLPDINSNRTAREFAERTAINTP IQGTAAADI IKLAMIKIHRIFKEKGFGTR	849
Tat	ETLFGKRKRPLPDINSNRTAREFAERTAINTP IQGTAAADI IKLAMIKLDTTIEEKGFEK	849
Thy	KTLAGRKRYIPEIFSPNKTVKELGQRIAVNTP IQGSASDLIKCAMVALYKELKKQNLKTS	791
Bst	TTLHRRRYLPDITSRNFNVRSAERTAMNTP IQGSAADI IKKAMIDLAAARLKEEQQLQAR	823
Gme	TTLHRRRYLPDITSRNFNVRSAERTAMNTP IQGSAADI IKKAMIDLAAARLKEEQQLQAR	823
Bcx	TTLHRRRYLPDITSRNFNVRSAERMAMNTP IQGSAADI IKKAMIDLNARLKEERLQAR	534
Gka	TTLHRRRYLPDITSRNFNVRSAERMAMNTP IQGSAADI IKKAMIDLNARLKEERLQAR	825
Afl	TTLHRRRYLPDITSGNFNVRSAERTAMNTP IQGSAADI IKKAMIDLANRLHEERLQTR	829
Bfs	TTLHRRRYIPELTSRNFNLRSAERTAMNTP IQGSAADI IKKAMIDMADKLKEKNLQAK	826
Toc	TTILNRRRYIPEINSRNYNLRSFAERVAMNTP IQGSAADI IKVAMVKITNHFREYGLKAK	811
Cfe	ETLFGRRRYIPELSSKNRTVQGFGER TAMNTP IQGSAADI IKLAMINVEKELKARKLKS	499
Taq	ETLFGRRRYVPDLEARVKSVRERAAERMAFNMPVQGTAAADLMKLAMVKLFPRLLEMG--AR	778
Mth	ETLFGRRRYIPDIRSRNRNVREAAERMAFNMPVQGTAAADLMKLAMVKLAPEIRSLG--AR	781

\* : \* : :\*:: : . : \* \* \* : \* : \* : \* : \* : \* : \* : \* : \* : \*



**Figure 3.13**

ClustalW alignment of the novel Pol I sequences identified in this project. The N-terminal, 5'-3' exonuclease, Fingers, Palm, Thumb and C-terminal subdomains are highlighted, as are common motifs. Critical residues, identified in Chapter 1, are highlighted in light blue (*B.st* Pol I numbering). The Singh *et al.* (2007) defined SYR residues, identified as essential to the strand-displacement activity of the *E.coli* Klenow Pol I, are highlighted in yellow (*E.co* Pol I numbering). The Large Fragment start positions have been defined using the reported *B.st* LF Pol I and *KlenTaq* Pol I positions (Green). The 3'-5' exonuclease key motifs (as defined by Derbyshire *et al.*, 1988) are highlighted in pink (*E.co* Pol I numbering)



**Figure 3.14**

A ClustalW (Thompson *et al.*, 1994) un-rooted phylogram of the LF DNA Pol I sequences providing a visual depiction of evolutionary relationships.

### 3.3 Discussion

Isothermal DNA amplification reactions are becoming an increasingly popular method for diagnostic level detection of diseases and pathogens. The majority of published methods rely solely on the mildly thermophilic *B.st* Large Fragment DNA Polymerase I enzyme to achieve the strand-displacement activity that isothermal amplification requires. A need for alternative strand-displacement DNA Polymerases, with potentially enhanced activities, was identified and is the basis of this study.

A variety of thermophilic organisms, optimally growing around 65°C, were chosen from which to identify novel DNA Pol I enzymes. The high growth temperature was selected with the aim to identify only enzymes thermostable and optimally active at the temperature required to provide highly specific target primer annealing. The LAMP method, identified for this study, requires six individual primers with eight specificities to enable target amplification. The 65°C LAMP reaction temperature defines the optimum temperature of the current enzyme of choice, *B.st* DNA Pol I, but also defines a high temperature for primer annealing and specificity. Too low a reaction temperature may introduce false priming, primer dimers, and lead to non-specific amplification products, potentially reducing the reaction efficiency. In contrast, too high a reaction temperature may be problematic to enable the eight primer sites to be designed with the required annealing temperature. A 65°C reaction was defined for this study as a compromise for ease of primer design whilst ensuring high primer specificity.

Organisms were identified for this report from a variety of sources. These included the DSMZ culture collection, donated thermophilic strains, and soil and water sampled directly from thermal features across the globe, ready to be screened for spore-forming thermophilic *Geobacillus*-like organisms. Twelve thermophilic organisms with growth temperature ranges between 55 to 80°C were identified and the genomic DNA successfully extracted.

A 16S rRNA comparison of the strains identified a diverse range of organisms, with those grouped close to *Geobacillus stearothermophilus* (*B.st*) Pol I expected to be the most promising candidates for a strand-displacing enzyme. Relying on 16S rRNA primers, designed using only those sequences from publically available databases, highlights the obvious limitation that organisms can only be identified if the employed primers are applicable to them. Under the current protocol of 16S rRNA identification, the whole bacterial world in environmental samples will remain partially invisible. However, for this study the universal primers designed by Lane *et al.* (1991) successfully enabled the

identification and thus taxonomic alignments of the organisms chosen for this study.

A complex gene-walking technique enabled the *polA* gene sequence to be revealed for those organisms not yet fully genome sequenced and publically available for analysis. The results in this chapter confirm the suitability of the reported Bergquist *et al.* (2004) CODEHOP primers to enable *polA* gene fragments to be identified without previous knowledge of the DNA sequence. The inverse PCR gene-walking method, identified by Triglia *et al.* (1989), remains an effective method to identify unknown *polA* flanking regions.

DNA Pol I amino acid sequence alignments using ClustalW (Thompson *et al.*, 1994) provides an indication of the homology between each enzyme and, more importantly, identifies the unique enzymes yet to be characterised. All *polA* sequences clearly identified the common Motif's: A, B, C, 1, 2, and 6 characterised by Delarue *et al.* (1990) in keeping with the common architecture and function of the DNA Pol I enzyme. The majority of DNA Pol I sequences show the expected architecture consisting of an N-terminal 5'-3' exonuclease domain, 3'-5' exonuclease domain (although not necessarily active), and 5'-3' DNA polymerase domain containing the Palm, Thumb and Fingers subdomains.

The sequence alignments enabled comparisons to be made with *T.aq* (*KlenTaq*) DNA Pol I (Korolev *et al.*, 1995) and *B.st* DNA Pol I (Kiefer *et al.*, 1997) and thus define the Large Fragment start positions for the closely related sequences (Figure 3.13). The 5'-3' exonuclease domain, identified in the majority of Pol I sequences, was not required for this study. The activity of this domain will inhibit the LAMP reaction through the degradation of the annealed looped primers by the Flap-endonuclease-like activity. The Large Fragment (Klenow or *KlenTaq*-like) polymerase was therefore used for this study. All further references to DNA Pol I in this thesis should therefore be assumed as being the Large Fragment, unless otherwise stated.

Several enzymes were suggested to possess an active 3'-5' exonuclease domain based on the sequence alignments (Figure 3.13). Two amino acid residues known to confer 3'-5' exonuclease activity in *E.co* (Derbyshire *et al.*, 1988) are present in the *K.ol* sequence, which confirms the taxonomic homology defined by the phylogenetic tree to *T.ma*. *T.in* and *T.at* DNA Pol I also show the required residues suggesting an active 3'-5' exonuclease domain. The activity of this domain provides the DNA Pol I with a proof-reading or editing function, desirable for many DNA amplification reactions, but not for LAMP. The 3'-5' exonuclease domain will potentially degrade the ssDNA oligonucleotide primers required for the reaction. This effect is expected to reduce the amplification speed due to the removal of

primers available to the reaction, and also contribute to non-specific amplification products due to false priming of partially degraded primers.

Singh *et al.* (2007) reported three essential residues to confer strand-displacement activity to *E.co* Klenow DNA Pol I. Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup> are highlighted (yellow) on the DNA Pol I alignment (Figure 3.13), as are the similar residues found within other DNA Pol I enzymes in the same position. DNA Pol I enzymes are involved with the removal of Okazaki fragments (Kornberg, 1980) and thus require strand-displacement activity to enable RNA primer removal. It is therefore not surprising to locate identical residues in the majority of the DNA Pol I sequences. Arg<sup>841</sup> is identical in all Pol I sequences, with Ser<sup>769</sup> present in all but the *T.ma* Pol I sequence. Singh *et al.* (2007) reported the presence of an aromatic residue (Phe, Trp, Tyr, His) at the apex of the Fingers domain enables DNA strand separation. All Pol I sequences possess an aromatic residue in this position. The Pol I sequences obtained indicate the ability to strand-displace and therefore imply a suitability as an alternative enzyme for isothermal DNA amplification reactions. Whether the *T.ma* DNA Pol I will be suitable will need to be assessed, due to the missing key Serine residue.

Having successfully identified twelve Large Fragment DNA Pol I sequences in this chapter, the *polA* genes will be cloned, and the recombinant protein over-expressed, to enable activity comparisons in a variety of assays including, importantly, the ability to strand-displace DNA.



## 4. CLONING, EXPRESSION & PURIFICATION

### 4.1 Introduction

#### 4.1.1 Polymerase genes identified

The results in Chapter 3 successfully identified twelve DNA *polA* genes from a wide variety of organisms. The strains vary extensively in their optimal growth conditions, with moderately thermophilic, thermophilic and extremely thermophilic organisms identified. A selection of *Bacillus* and *Geobacillus* organisms were purposefully chosen for this study with the knowledge that *Geobacillus stearothermophilus* encodes a DNA Pol I (*B.st* LF DNA Pol I) with sufficient strand-displacement activity for isothermal amplification reactions. *Bacillus subtilis* DNA Pol I (NEB, UK) and *Bacillus smithii* DNA Pol I (Thermo Scientific, UK) are further examples of commercially-available strand-displacing DNA polymerase enzymes, but their lower optimal reaction temperatures reflect the bacteria's reduced growth temperatures (30°C and 55°C, respectively). The genera *Bacillus* and *Geobacillus* include a wide variety of bacteria with a broad range of growth conditions including temperature, pH and salt conditions for optimal growth (Nazina *et al.* 2001). It was proposed that additional isolates from these genera may encode alternative polymerases with strand-displacement activity.

Sequence alignments in Chapter 3 confirmed the shared regions and motifs expected from all DNA Pol I enzymes, as reported by Delarue *et al.* (1990), indicating the conserved nature of the DNA repair function of this group of enzymes. *Carboxydotherrnus ferrireducens* (*C.fe*) DNA Pol I shares 99% identity with the previously reported *Carboxydotherrnus hydrogenoformans* (*C.hy*) DNA Pol I, shown to possess RNA-dependent DNA polymerase activity suitable for RT-PCR reactions (Angerer *et al.*, 1998). No examples can be found in the literature demonstrating the enzyme's ability to strand-displace DNA and this will need to be investigated further.

Three of the Pol I enzymes identified in this thesis possess the two key residues (Derbyshire *et al.*, 1988) necessary to confer an active 3'-5' exonuclease domain, incorporating a proof-reading mechanism ability to *T.in*, *T.at* and *K.ol* Pol I enzymes. Whether these enzymes exhibit this activity, and are suitable for a LAMP reaction, is to be investigated in this chapter.

The FLAP endonuclease (FEN) or 5'-3' exonuclease domain, if present, confers 5' nuclease activity to the DNA Pol I enzyme and plays a major role in processing the RNA primers that are used to initiate lagging-strand (Okazaki fragment) synthesis *in vivo*. A FEN can cleave the 5'-flap structures one nucleotide into the double-stranded region immediately downstream of a single-stranded 5' arm. They also degrade free 5' ends of single or double

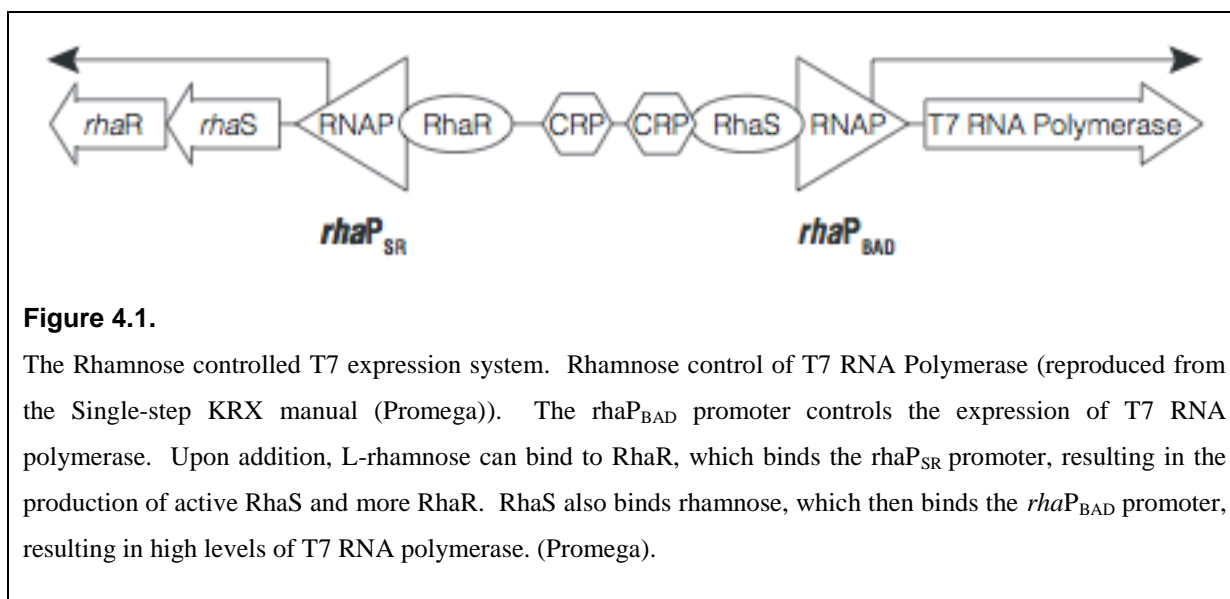
stranded DNA through their exonucleolytic activity (Joyce and Steitz, 1987). These activities are useful in biotechnology for certain applications, e.g. Taqman probe PCR assays. FEN activity is not desirable for many DNA amplification methods and therefore the Large Fragment, *KlenTaq*/Klenow-like DNA Pol I enzymes, missing the 5' nuclease domain, are preferred. The LAMP method relies on the self-complementary looping of primers to the target DNA, exposing potential sites for FENs to cleave. As such, only the Large Fragment Pol I will be suitable for this isothermal amplification method and only these were therefore cloned for this study. All references to Pol I should therefore be assumed to be the Large Fragment (LF) variant, unless otherwise stated.

With a wide variety of different Pol I enzymes identified, it was hoped that some would harbour sufficient strand-displacement activity, enabling their use as alternative enzymes to *B.st* DNA Pol I for use in isothermal DNA amplification methods, specifically LAMP. To identify this activity the *polA* genes would need to be cloned and the recombinant protein overexpressed, to provide active and soluble DNA Pol I for purification.

#### **4.1.2 Recombinant cloning and expression**

Recombinant DNA cloning is widely used in biotechnology and research and can be applied to the over-expression of foreign prokaryotic proteins in *E.coli* strains. The pET<sup>®</sup> system is one such vector system for cloning (Novagen, USA) and requires genes to be cloned into plasmids under the control of strong bacteriophage T7 transcription. The expression of protein is induced by providing a source of T7 RNA polymerase in the host cell. The pET<sup>®</sup> system importantly offers the ability to maintain target genes transcriptionally silent in the un-induced state, enabling the controlled expression of protein potentially toxic to the host cell (Novagen pET<sup>®</sup> System manual, 10<sup>th</sup> edition).

The pET<sup>®</sup> vector requires a compatible *E.coli* host strain. KRX (Promega, USA) is an *E.coli* K12 derivative with attributes enabling its use as a combined cloning and expression host (KRX manual, 2010). The strain has been engineered to lack the most common *E.coli* nuclease (*endA*), and the *recA*- mutation to minimise undesirable recombination events (KRX manual). KRX also contains mutations to reduce a source of proteolysis (*ompT*- and *ompP*-) of over-expressed proteins in *E.coli*. Importantly, KRX incorporates a chromosomal copy of the T7 RNA polymerase gene enabling its use in the T7 expression system. Uniquely, the T7 RNA polymerase gene is driven by a strong rhamnose promoter (*rhaP*<sub>BAD</sub>), to control recombinant protein expression. Figure 4.1. depicts the rhamnose control of T7 RNA polymerase expression in KRX.



Cloning a foreign gene into an *E.coli* host can have limitations for its recombinant protein expression. Not all of the mRNA codons are used equally in all proteins and often vary by organism. The frequency of the codon usage is reflected by the abundance of the organism's common tRNAs. Major codons are those expressed in highly expressed genes, whereas the minor expressed codons tend to be in genes expressed at a lower level within the cell. This study requires the cloning of a diverse range of *polA* genes from a wide range of organisms. The broad codon usage potentially required may not be reflected by the *E.coli* KRX host organism. This may lead to problems during expression, including interrupted translation, frame shifting mis-incorporation of amino acids, or inhibition of protein synthesis. This may lead to reduced or no expression of the recombinant protein (Novy *et al.*, 2001). To avoid this potential problem, methods have been developed to increase the rare tRNA genes available to the *E.coli* host. *E.coli* Rosetta<sup>TM</sup>2 (Novagen, UK) contains a pRARE2 plasmid that encodes 7 minor tRNAs in *E.coli*: AUA (isoleucine), AGG/AGA and CGG (arginine), CUA (leucine), CCC (proline), and GGA (glycine). The addition of pRARE2, providing these rare tRNAs, into *E.coli* KRX was therefore expected to aid expression of the DNA Pol I enzymes in this study.

#### 4.1.3 Purification

To accurately assess the activity of the Pol I enzymes, it will be necessary to purify the enzymes away from background *E.coli* proteins. Enzyme activity characterised directly from cell extracts is likely to contain contaminating DNases and RNases. DNases may degrade the template required for a LAMP reaction and additional downstream DNA-based characterisation assays. RNases may degrade the RNA template required for

complementary DNA (cDNA) synthesis. A heat treatment step can be applied directly to the cell extract to remove the majority of DNases and RNases. For example, commercially available *E.coli* DNase I is denatured by heating at 70°C for 10min (NEB). *E.coli* RNaseHI is reported to denature above 50°C (Kanaya et al., 1996). Purification of the DNA Pol I away from these contaminants will therefore be necessary before enzyme characterisation.

Several different purification approaches are available to overcome the potential problems mentioned above.

#### Affinity column chromatography

Recombinant proteins can be expressed as fusions with short affinity tags such as a polyhistidine (6x His) tag. The addition of a 6x His tag at the N- or C-terminus of the DNA Pol I sequence permits immobilised metal affinity chromatography (IMAC) purification. HisPur™ Ni-NTA chromatography columns (Thermo Scientific, UK) were chosen for this study due to their superior binding affinities (60mg/ml of resin). The resin is composed of nickel-charged ( $\text{Ni}^{2+}$ ) nitrotriacetic acid (NTA) chelate immobilised onto 6% cross-linked agarose (Thermo Scientific). An exposed His-tag, on the surface of the protein, will bind to the Ni-NTA column, with non-tagged proteins passing straight through. An imidazole gradient, which has a higher affinity to the  $\text{Ni}^{2+}$  than the Histidine tag, permits the elution of the bound, purified protein from the column. Due to their hydrophilic and flexible nature, His-tags have also been shown to increase the solubility of target proteins and rarely interfere with the protein's function (GE Healthcare, UK).

#### Heparin Sepharose affinity column chromatography

The ligand heparin mimics the similarly polyanionic structure of nucleic acid and acts as an affinity ligand and a cation exchanger, enabling nucleic acid binding proteins to interact and reversibly bind (GE Healthcare, UK). The elution of bound proteins is carried out with a salt gradient that can potentially separate the DNA Pol I from other nucleic acid binding proteins depending on their affinity to the column.

#### Ion exchange column chromatography

Ion exchange chromatography separates proteins on the basis of differences in their net surface charge in relation to the pH of their environment. Anion exchange resins have a positive charge and are used to separate negatively charged compounds, while cation exchange resins offer a negative charge to separate positively charged molecules. The isoelectric points of the DNA Pol I proteins, reported in Chapter 3, define which ion exchange

column will be used for purification. The majority of Pol I enzymes show a pI value between 5.0-6.8 enabling purification with a ResourceQ anion exchange column using a buffer pH above the pI value. This purification step enables final polishing of the purified Pol I enzymes.

#### **4.1.4 Protein quantification**

The purified protein will require quantification to enable accurate downstream enzyme comparisons. The Qubit<sup>®</sup> protein assay (Invitrogen, UK) uses a fluorescent dye to determine the concentration of protein in a sample. The dye's fluorescent intensity increases when bound specifically to protein and the amount of fluorescence signal is directly proportional to the concentration of protein in the solution. The fluorescence signal can be directly converted to a protein concentration using standards of known concentration. The Qubit<sup>®</sup> assay has been selected for this study because it offers advantages over traditional UV absorbance based protein calculation methods (e.g. Lowry assay or Bradford assay) in that it is not affected by the presence of background DNA or RNA contaminants and is highly specific to protein. The assay permits sensitive and accurate detection of protein well below the limit of UV detection methods and across a wider pH range. In contrast, UV detection methods indiscriminately measure anything that absorbs at 280nm, including degraded nucleic acids, and free nucleotides.

#### **4.1.5 Chapter overview**

This chapter reports the detailed steps taken to clone, overexpress and purify the (LF) DNA Pol I from *G.me* and *T.at*. The *polA* genes were over-expressed in *E.coli* (KRX\_pRARE2) and purified to homogeneity using column chromatography. The remaining DNA Pol I were cloned, expressed and purified in an identical manner, and as such, are only summarised in this chapter to prevent repetition.

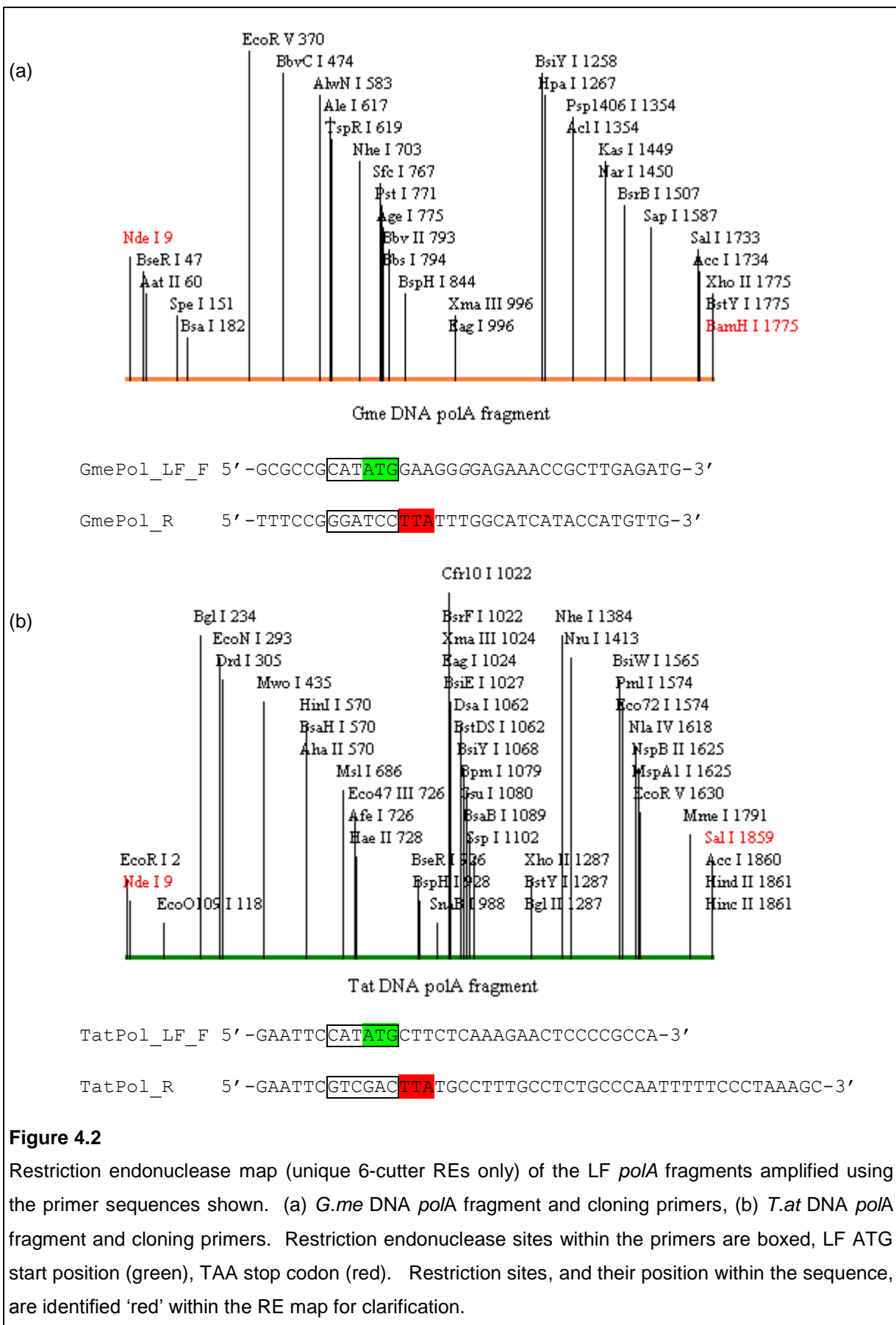
A further two DNA Pol I clones were kindly donated by Dr Clark (GeneSys Biotech Ltd.) for use in this study. The Large Fragment of *Thermus aquaticus* Pol I (*T.aq* Pol I) and the Large Fragment of *Thermotoga maritima* Pol I (*Tma* Pol I), mutated to inactivate 3'-5' exonuclease activity, were received as a plasmid DNA (pET24a+HIS\_*polA*) stock. Neither enzyme has been reported to show significant strand-displacement activity, and therefore will be suitable for subsequent enzyme comparisons.

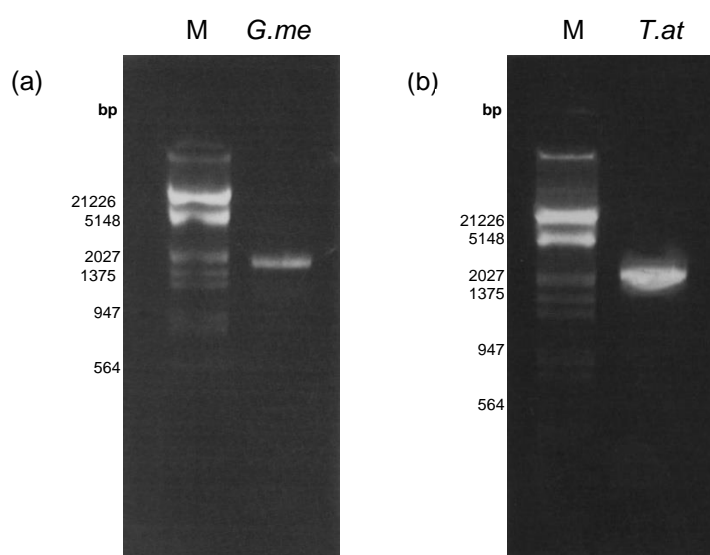
## 4.2 Results

### 4.2.1 Cloning of *G.me* and *T.at* LF DNA *polA*

Chapter 3 identified the Large Fragment ATG start positions of the twelve Pol I enzymes, with assignments based on sequence alignments with *KlenTaq* and *B.st* LF Pol I. Specific cloning primers were designed against the 1764bp *G.me* LF *polA* and 1851bp *T.at* LF *polA* sequences (Appendix i.iv). Primers contained an *Nde* I restriction endonuclease site immediately upstream, incorporating the ATG start position, and a *Bam*H I or *Sal* I restriction endonuclease site immediately downstream of the TAA stop codon (Figure 4.2). The primers enabled directional cloning into the pET24a+HIS vector as described in Chapter 2.3.1. The wild-type (WT) *polA* sequences were shown to contain a weak opal (TGA) stop codon (Vlasov *et al.*, 2012). The codon was changed to the more efficient ochre (TAA) codon by introducing the 'TAA' sequence into the reverse primer.

*G.me* LF *polA* and *T.at* LF *polA* were PCR amplified using a high fidelity Phusion™ DNA polymerase (Figure 4.3). Amplified products were visualised on an agarose gel to confirm the expected sized band had been amplified (Figure 4.3). The PCR products were purified and double-digested with *Nde* I and *Bam*H I or *Sal* I (NEB, UK). Purified and digested PCR products were ligated into pET24a+HIS. The vector had been previously digested using the same REs, and treated with Antarctic Phosphatase (NEB, UK) to remove the 5'-phosphate group, preventing self-ligation. The ligation reactions were transformed into *E.coli* KRX (Promega, UK), containing the additional chloramphenicol-resistant pRARE2 vector, and plated on LB plates with both *kan* and *Cr* for selection. Colonies were screened by PCR and recombinant clones verified by sequencing using the vector specific T7 primers (see Appendix i.iii for primer sequences).





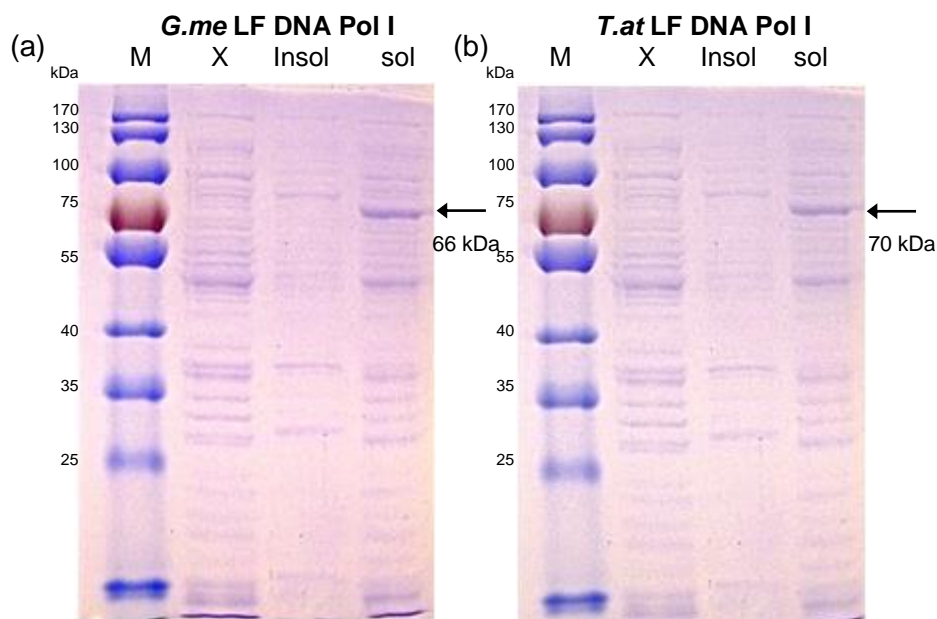
**Figure 4.3**

A 1.5% (w/v) agarose gel showing 5 $\mu$ l *polA* PCR product (a) 1785bp high fidelity PCR amplified fragment from *G.me* gDNA, (b) 1872bp high fidelity PCR amplified fragment from *T.at* gDNA. M (500 $\mu$ g *EcoR* I/*Hind* II DNA ladder).

#### 4.2.2 Small-scale expression of *G.me* and *T.at* LF DNA Polymerase I

A single recombinant colony of each Pol I was selected for protein expression studies and inoculated into an overnight culture medium. 10ml TB (*Kan*) was inoculated with 100 $\mu$ l of the overnight culture and incubated at 37°C with aeration at 275rpm in baffled shakeflasks following the method detailed in Chapter 2.4.1. At an OD<sub>600</sub> 0.8 the temperature of incubation was reduced to 24°C, as recommended by the KRX induction protocol, to aid protein solubility during induction. On reaching an OD<sub>600</sub> of 1, the culture was induced with L-rhamnose and IPTG. The cultures were further incubated overnight (approximately 22h) at 24°C with aeration at 275rpm. The cells were harvested by centrifugation, and the pellet was resuspended in cell lysis buffer and sonicated to break open the cells to release the soluble protein into the supernatant. The sample was centrifuged to pellet the cell debris, with the supernatant retained as the soluble fraction. The cell pellet was washed, resuspended in cell lysis buffer and a sample used as the insoluble fraction. 100 $\mu$ l culture equivalent samples of these fractions were electrophoresed on a 12% SDS-PAGE gel to confirm successful expression and to assess the solubility of the expressed protein. The gels showed 100% soluble expression of the LF DNA Pol I enzymes with a protein of approximately 65-70kDa (Figure 4.4).





**Figure 4.4**

12% SDS PAGE showing small scale protein expression. (a) 100 $\mu$ l culture equivalent *G.me* LF Pol I, (b) 100 $\mu$ l culture equivalent *T.at* LF Pol I. Lane 1: PageRuler Protein Ladder (M), Lane 2: uninduced (X), Lane 3: Insoluble fraction (Insol), Lane 4: Soluble fraction (sol). The Pol I expressed protein band is identified by an arrow.

#### 4.2.3 Remaining LF DNA polymerase I cloning and expression

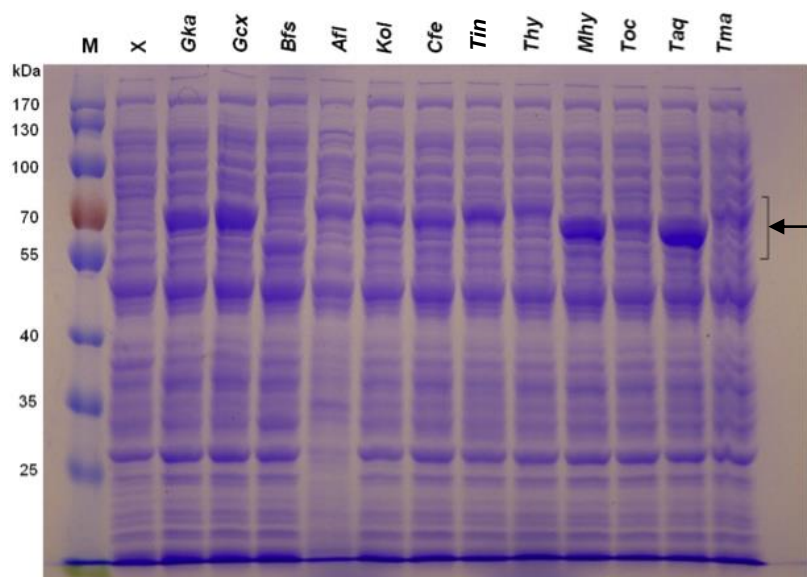
Only one *polA* gene, that from *Anoxybacillus flavithermus* (*A.fl*), was found to contain internal *Nde* I recognition sites within the LF DNA sequence. This prevented the use of the pET24a+HIS vector for directional cloning. *A.fl polA* was therefore cloned into pET24d+HIS, which permitted the use of *Nco* I (C/CATGG). To ensure the correct open reading frame of the gene, an additional glycine residue (Gly/G) was incorporated after the Large Fragment start, allowing the ATG of the *Nco* I restriction site to be used as the start codon. All *Nco* I directionally cloned *polA* genes would therefore start 'ATGGGA'. The additional amino-acid residue was not expected to alter the protein's function or solubility.

Table 4.1 details the restriction sites used to directionally clone each LF *polA* gene into the appropriate pET<sup>®</sup> vector. Cloning and over-expression of additional DNA Pol I enzymes were carried out in a similar fashion to that reported for the *G.me* and *T.at* DNA Pol I. All polymerases were shown to be 100% soluble and over-expressed with varying protein yields. An intense, overexpressed protein band at the expected molecular weight for each Pol I could be seen when analysed on a 12% SDS PAGE gel (Figure 4.5).

Strain	LF <i>polA</i> (bp)	LF Pol I (aa)	RE's used to clone	Isoelectric point (pI)	Calculated molecular weight (kDa)
<i>G.me</i>	1764	587	<i>Nde</i> I/ <i>Bam</i> H I	5.25	66.66
<i>T.at</i>	1848	615	<i>Nde</i> I/ <i>Sal</i> I	6.77	70.31
<i>G.ka</i>	1764	587	<i>Nde</i> I/ <i>Bam</i> H I	5.58	66.96
<i>A.fl</i>	1764	587	<i>Nco</i> I/ <i>Bam</i> H I	6.04	66.98
<i>K.ol</i>	1851	616	<i>Nde</i> I/ <i>Sal</i> I	5.73	69.98
<i>M.hy</i>	1659	552	<i>Nde</i> I/ <i>Sal</i> I	6.70	62.20
<i>B.fs</i>	1764	587	<i>Nde</i> I/ <i>Bam</i> H I	5.37	66.46
<i>G.cx</i>	1764	587	<i>Nde</i> I/ <i>Bam</i> H I	5.64	66.95
<i>C.fe</i>	1659	552	<i>Nde</i> I/ <i>Sal</i> I	5.71	63.12
<i>T.in</i>	1851	615	<i>Nde</i> I/ <i>Sal</i> I	6.77	70.20
<i>T.hy</i>	1704	567	<i>Nde</i> I/ <i>Sal</i> I	8.73	65.58
<i>T.oc</i>	1764	587	<i>Nde</i> I/ <i>Bam</i> H I	5.88	66.23
<i>T.aq</i>	1662	553	<i>Nde</i> I/ <i>Sal</i> I	5.86	62.35
<i>T.ma</i>	1833	610	<i>Nde</i> I/ <i>Sal</i> I	5.59	66.69

**Table 4.1**

The isoelectric point (pI) was estimated using the MBCS programme using the reported LF DNA Pol I sequence. All sizes are those excluding the 6x Histidine N-terminal tag (predicted to add 0.84kDa to the protein).



**Figure 4.5**

12% SDS PAGE comparing small-scale protein expression of 100μl culture equivalent soluble protein from whole cell homogenate. Lane 1: PageRuler Protein Ladder (M), Lane 2: uninduced sample (X), The arrow highlights the overexpressed protein in each well.

The small-scale expression study successfully demonstrated all DNA Pol I clones to express soluble enzyme. The small scale expression study had confirmed:

- (i) The T7 expression system had successfully over-expressed the recombinant protein.
- (ii) All *polA* genes were cloned in the correct reading frame.
- (iii) All TAA stop codons had prevented translational read-through.
- (iv) The 24°C induction temperature resulted in 100% soluble protein.

Additional protein was required to enable the recovery of enzyme after several rounds of purification by column chromatography. A 100ml (large-scale) culture of each LF DNA Pol I was over-expressed following the methods described in Chapter 2.4.2 to provide pure enzyme for characterisation.

### 4.3 Purification

DNA polymerase activity could not be accurately determined using protein directly from unfractionated cell extracts due to the possible presence of bacterial DNases and inhibiting *E.coli* proteins. DNases may degrade the template DNA inhibiting characterisation assays. It was therefore necessary to purify each DNA Pol I before assays could be performed. To save repetition, the purification of *G.me* and *T.at* LF DNA Pol I enzymes are detailed below, following the methods described in Chapter 2.5, with the remaining enzymes summarised later in the chapter.

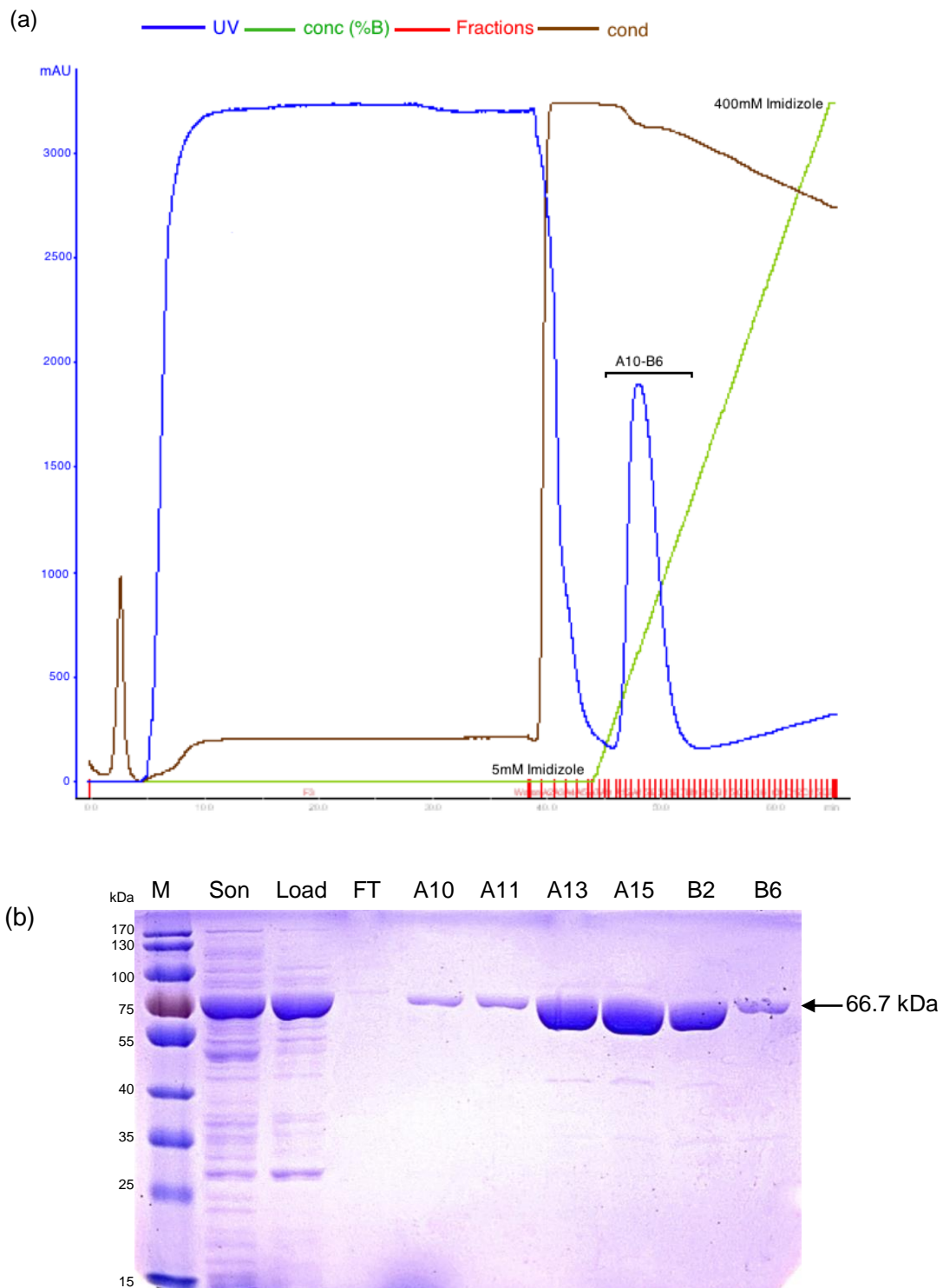
#### 4.3.1 Purification of *G.me* and *T.at* LF DNA Polymerase I

The recombinant clones pET24a+HIS\_*Gme* LF DNA Pol I and pET24a+HIS\_*Tat* LF DNA Pol I in *E.coli* KRX(pRARE2) were selected for large-scale expression following the methods detailed in Chapter 2.4.2. A 2.5ml LB (*Kan/Chr*) culture was grown up overnight and 1ml inoculated into a 100ml TB culture (*Kan*). Cultures were induced as before and the cells harvested after an overnight induction at 24°C. The overnight OD<sub>600</sub> was recorded at 18.6 for the *G.me* induced sample, and 19.2 for the *T.at* induced sample. The un-induced control gave a reading of 20.2; therefore, the recombinant proteins were shown to have limited effect on the *E.coli* growth. Cells were resuspended in 10ml Purification Buffer A and sonicated on ice. Samples were heated at 65°C for 60min to denature background *E.coli* proteins, cooled on ice and then centrifuged to remove cell debris. The supernatant was

retained and brought to 20ml with Purification Buffer A ready to be loaded onto a 1ml Ni-NTA HisPur™ column.

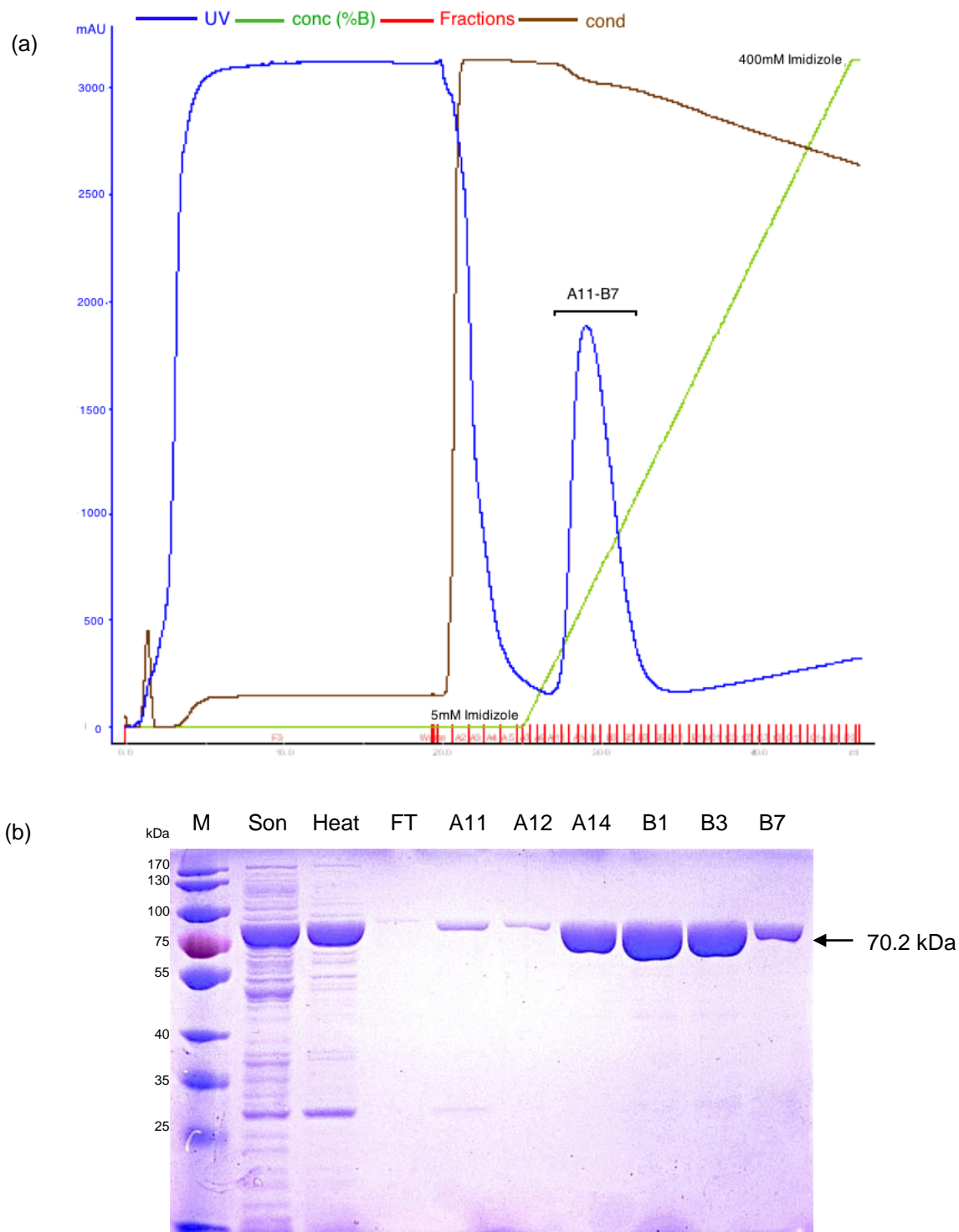
#### Metal chelating affinity chromatography

The 20ml samples were loaded onto a pre-equilibrated 1ml Ni-NTA HisPur™ column. The column was washed with Purification Buffer A, then protein was eluted with a 10CV gradient of Purification Buffer B (5-400mM imidazole), collecting 1ml fractions. Peak fractions, confirmed by electrophoresing a 5µl sample on a 12% SDS-PAGE gel, were pooled and dialysed overnight against Purification Buffer C to reduce the NaCl concentration, ready for loading onto the next purification column. Figure 4.6 and 4.7 detail the His-tagged purification results of *G.me* and *T.at* LF DNA Pol I enzymes, respectively.



**Figure. 4.6**

*G.me* LF DNA Pol I 1ml HisPur NiNTA column chromatography. (a) Purification chromatogram: Elution peak (9-36% B, peak at 19% B). (b) 12% SDS-PAGE gel comparing purification fractions. PageRuler protein ladder (M), 10 $\mu$ l sonicated sample (Son), (10 $\mu$ l) heat-treated sample/Column loading sample (Load), 10 $\mu$ l column flow through (FT), 5 $\mu$ l elution fractions A10-B6. An arrow identifies the overexpressed *G.me* Pol I.



**Figure 4.7**

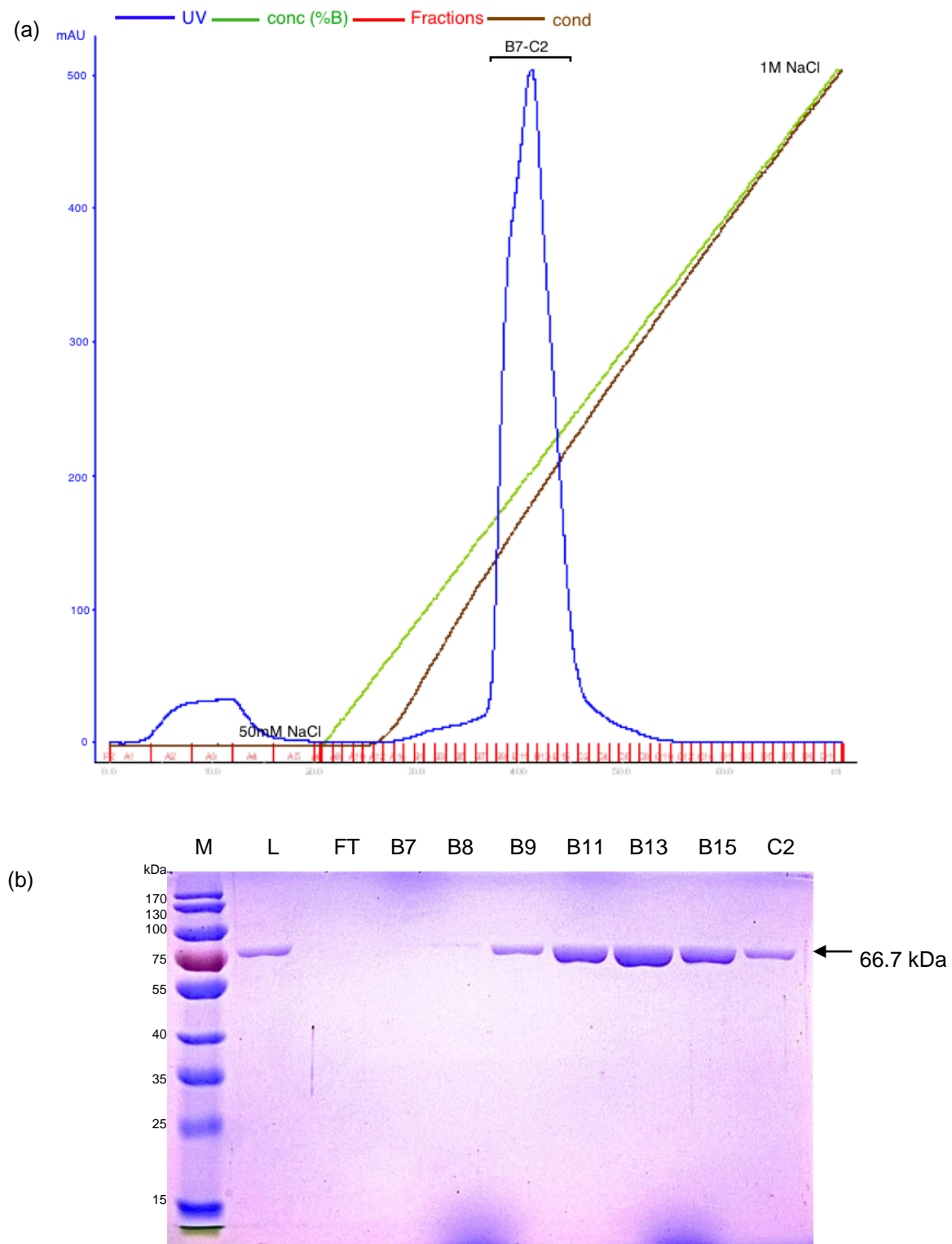
*T.at* LF DNA Pol I 1ml HisPur NiNTA column chromatography. (a) Purification chromatogram: Elution peak (10-37% B, peak at 19.5% B). (b) 12% SDS-PAGE gel comparing purification fractions. PageRuler protein ladder (M), 10 $\mu$ l sonicated sample (Son), (10 $\mu$ l) heat-treated sample/Column loading sample (Load), 10 $\mu$ l column flow through (FT), 5 $\mu$ l elution fractions A11-B7. An arrow identifies the overexpressed *T.at* Pol I.

### Heparin Sepharose column purification

Heparin Sepharose column chromatography was used as a second step for purification of the DNA Pol I enzymes. The heparin ligand was used to take advantage of the nucleic acid binding properties of the DNA Pol I enzymes to purify them from the background *E.coli* proteins. The dialysed His-tagged purified protein fractions were brought to 20ml using Purification Buffer C (low salt) and then loaded onto a 5ml HiTrap™ Heparin Sepharose column, pre-equilibrated in Purification Buffer C. The column was further washed with Purification Buffer C to remove unbound protein, and then eluted with Purification Buffer D (0.05-1M NaCl). Peak fractions, confirmed by electrophoresing a 5µl sample on a 12% SDS-PAGE gel, were pooled and fully dialysed against Purification Buffer C to reduce the NaCl concentration, ready for loading onto the next purification column. Figure 4.8 and 4.9 detail the Heparin Sepharose purification results of *G.me* and *T.at* LF DNA Pol I enzymes.

### Ion Exchange column purification

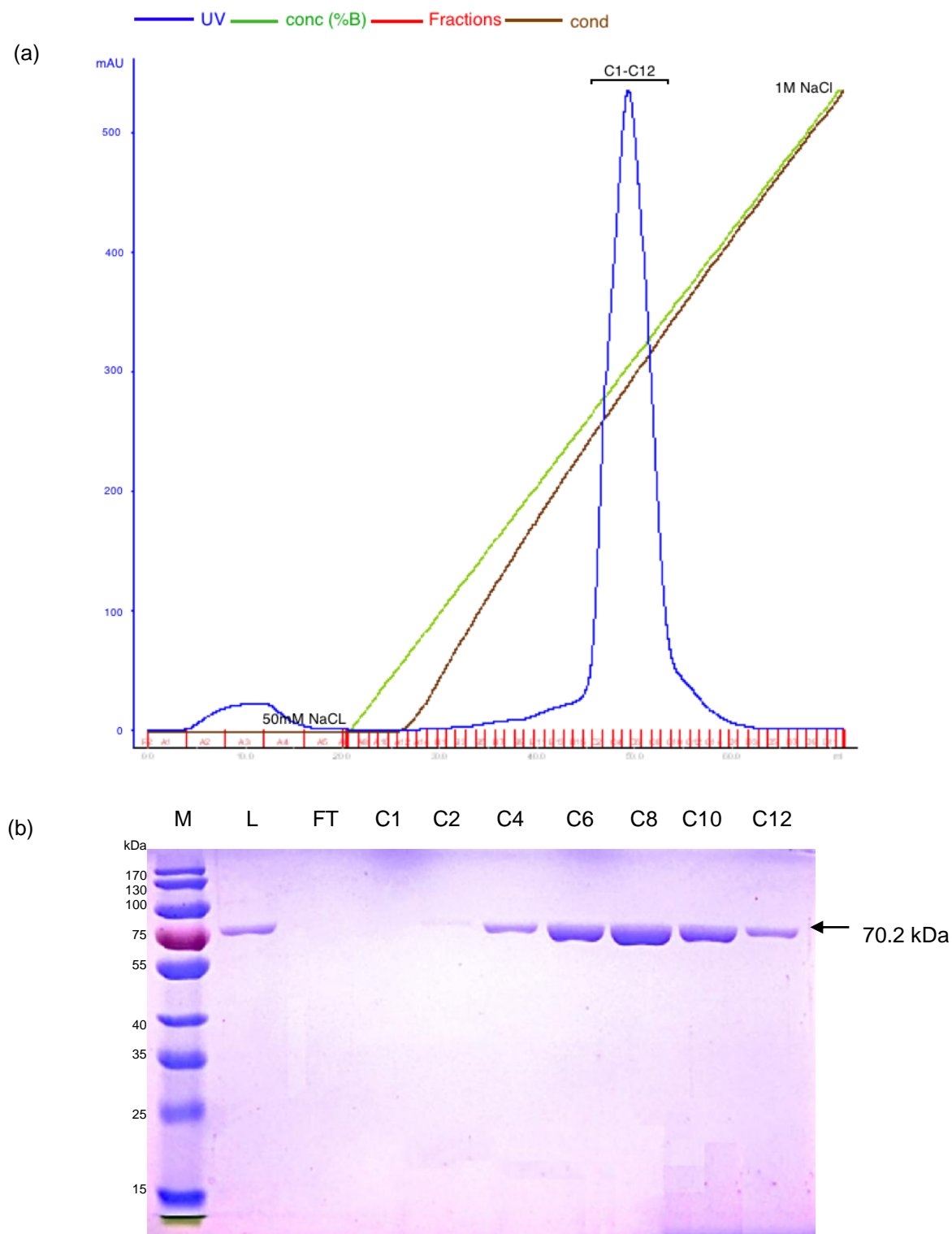
A final polishing step of the DNA Pol I enzymes was achieved using 1ml ion exchange column chromatography. The volume of the dialysed samples from the heparin purification was increased to 20ml using Purification Buffer C (low salt) and loaded onto a 1ml ResourceQ or ResourceS column (GE Healthcare, UK) depending on the pI of the protein to be purified. The pI of *G.me* and *T.at* LF DNA Pol I were calculated from the amino acid composition and determined to be 5.25 and 6.77 respectively (Table 4.1). The proteins were therefore purified using the anion exchange ResourceQ column using Purification Buffer C (pH8.0). The samples were loaded and the column was washed with 5 column volumes of Purification Buffer C. The proteins were finally eluted with a 10CV gradient of Purification Buffer D (0.05-1M NaCl). Peak fractions were confirmed by electrophoresing 5µl samples on a 12% SDS-PAGE gel. Figure 4.10 and 4.11 detail the ResourceQ purification results of *G.me* and *T.at* LF DNA Pol I enzymes. Peak fractions were pooled and fully dialysed against a 50% (v/v) glycerol-containing, detergent-free, Polymerase Storage Buffer. 100µl aliquots of the two column purified DNA Pol I enzymes were transferred to fresh tubes to enable characterisation of the enzyme, in reactions where detergents may inhibit particular assay components. Concentrated detergent stocks were added to the remaining enzyme sample for long term storage at -20°C.



**Figure 4.8**

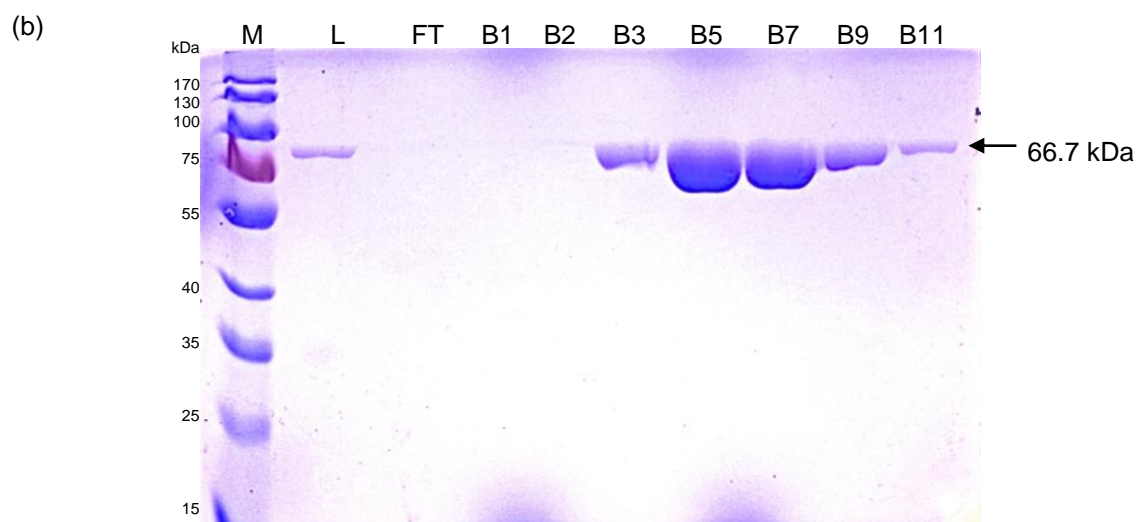
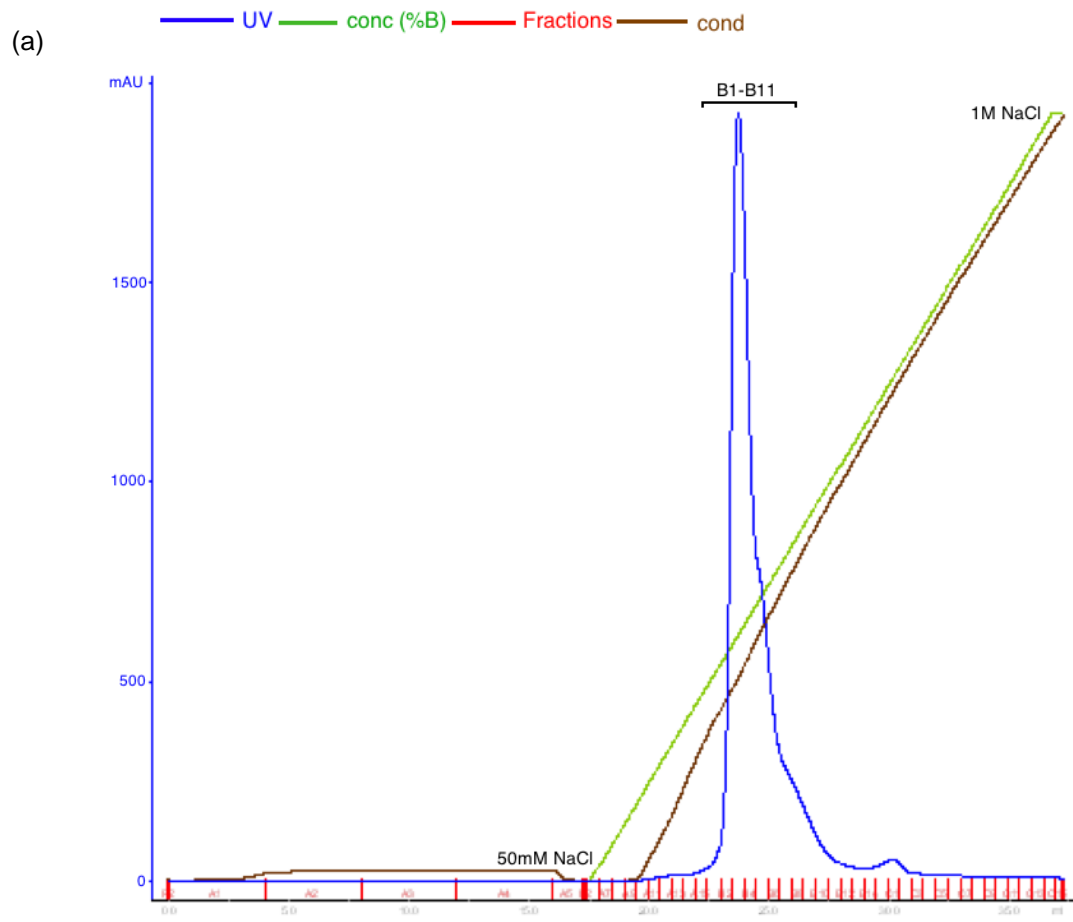
*G.me* LF DNA Pol I 5ml HiTrap Heparin Sepharose 6FF chromatography. (a) Purification chromatogram: Elution peak (32-50% B, peak at 40% B). (b) 12% SDS-PAGE gel comparing purification fractions. PageRuler protein ladder (M), 10 µl column load (L), 10 µl column flow through (FT), 5 µl elution fractions B7-C2. An arrow identifies the overexpressed *G.me* Pol I.





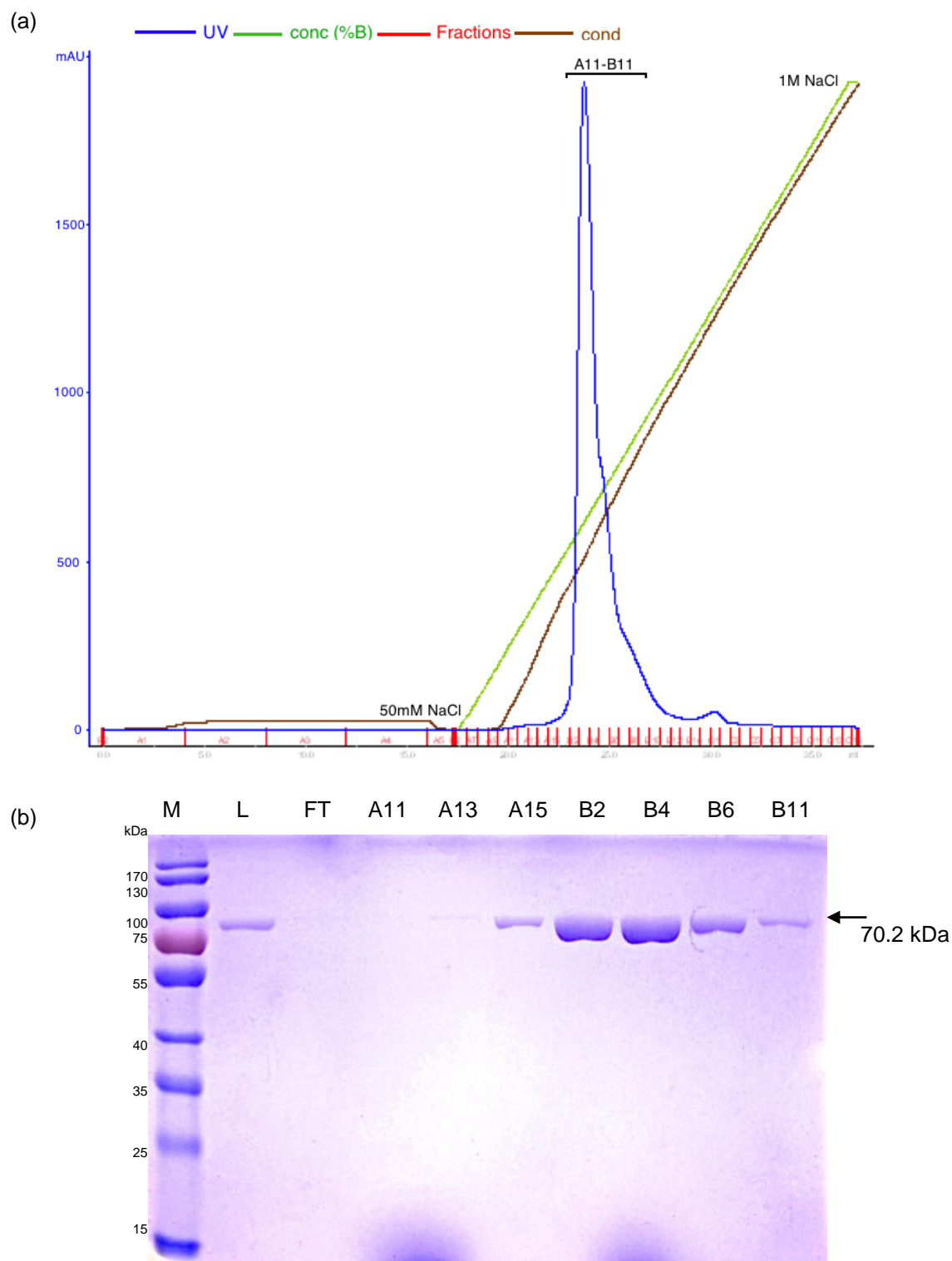
**Figure 4.9**

*T.at* LF DNA Pol I 5ml HiTrap Heparin Sepharoe 6FF chromatography. (a) Purification chromatogram: Elution peak (48-68% B, peak at 57% B). (b) 12% SDS-PAGE gel comparing purification fractions. PageRuler protein ladder (M), 10 $\mu$ l column load (L), 10 $\mu$ l column flow through (FT), 5 $\mu$ l elution fractions C1-C12. An arrow identifies the overexpressed *T.at* Pol I.



**Figure 4.10**

*G.me* LF DNA Pol I 1ml ResourceQ ion-exchange chromatography. (a) Purification chromatogram: Elution peak (28-48% B, peak at 32% B). (b) 12% SDS-PAGE gel comparing purification fractions. PageRuler protein ladder (M), 10 $\mu$ l column load (L), 10 $\mu$ l column flow through (FT), 5 $\mu$ l elution fractions B1-B11. An arrow identifies the overexpressed *G.me* Pol I.

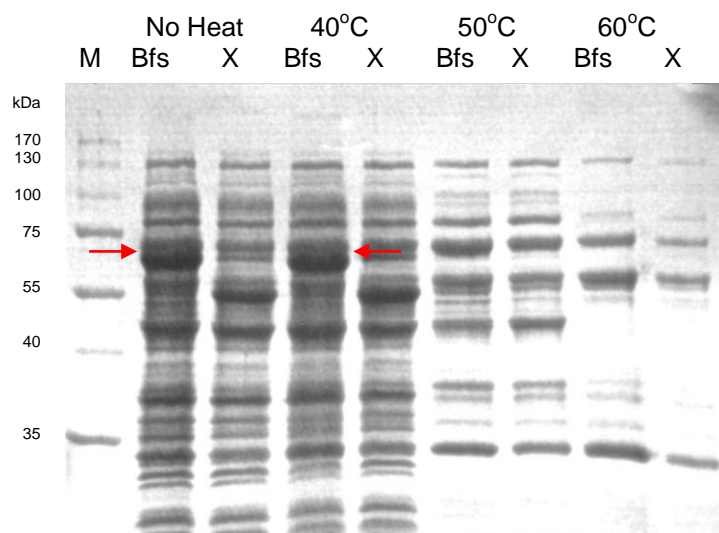


**Figure 4.11**

*T.at* LF DNA Pol I 1ml ResourceQ ion-exchange chromatography. (a) Purification chromatogram: Elution peak (8-19% B, peak at 13.5% B). (b) 12% SDS-PAGE gel comparing purification fractions. PageRuler protein ladder (M), 10μl column load (L), 10μl column flow through (FT), 5μl elution fractions A11-B11. An arrow identifies the overexpressed *T.at* Pol I.

### 4.3.2 Purification of remaining LF DNA polymerase I

The thermostability of *Bacillus furnas* (*B.fs*) LF DNA Pol I was assessed before purification of the enzyme on a larger scale. This was because *Bacillus furnas* could only be revived at 50°C in Chapter 3. Its encoded proteins were therefore not expected to be significantly more thermostable. Aliquots from the small scale *B.fs* expression culture were incubated at 40°C, 50°C and 60°C in a waterbath for 60 min to ascertain whether the enzyme would be suitable for further investigation. 100µl culture equivalent samples were loaded onto a 12% SDS-PAGE gel to compare the effects of heat treatment on the enzyme (Figure 4.12). *B.fs* LF Pol I was shown to be fully denatured during the 50°C heat step and as such was removed from this study due to its lack of thermostability.



**Figure 4.12**

12% SDS-PAGE gel comparing the effects of 60min heat denaturation steps (40/50/60°C) on *B.fs* LF DNA Pol I. PageRuler protein ladder (M), induced sample (*B.fs*), uninduced control sample (X). The induced protein band, identified by the red arrows at approximately 65kDa, is completely denatured at 50°C.

The remaining DNA Pol I enzymes were purified following the same method used for *G.me* and *T.at* LF DNA Pol I. Table 4.2 details the individual large-scale growth characteristics and expected kDa of each LF Pol I. One notable difference in the purification procedure was for the purification of *T.hy* LF DNA Pol I. Showing a pI value of 8.73 (Table 4.1), which is above the pH of the Purification Buffer C (pH8.0), a 1ml cation exchange ResourceS (GE

Healthcare, UK) column was used for the ion exchange purification step. The literature suggests the pI should be at least one pH unit above the buffer pH to enable efficient purification (GE Healthcare – Resource Q manual). However, purification proved to be successful using the existing Purification Buffer C, with no visible Pol I protein in the flow through fraction.

Strain	100ml culture OD <sub>600</sub>	100ml cells wet weight (g)	Molecular weight (kDa)
<i>G.me</i>	18.6	3.6	66.66
<i>T.in</i>	18.9	3.7	70.20
<i>G.ka</i>	18.4	3.6	66.96
<i>A.fl</i>	18.8	3.6	66.98
<i>K.ol</i>	17.9	3.5	69.98
<i>M.hy</i>	19.3	3.8	62.20
<i>G.cx</i>	20.3	4.0	66.95
<i>C.fe</i>	19.0	3.8	63.12
<i>T.at</i>	19.2	3.9	70.31
<i>T.hy</i>	20.1	4.0	65.58
<i>T.oc</i>	18.5	3.6	66.23
<i>T.aq</i>	18.5	3.6	62.35
<i>T.ma</i>	17.8	3.5	66.69

**Table 4.2**

Large-scale protein expression comparison data. OD<sub>600</sub> taken after 22h induction at 24°C. An uninduced control showed an OD<sub>600</sub> 20.2 with a wet cell weight of 4g. The predicted molecular weight (kDa) of each protein was calculated using the MBCS programme.

#### 4.3.3 Protein quantification

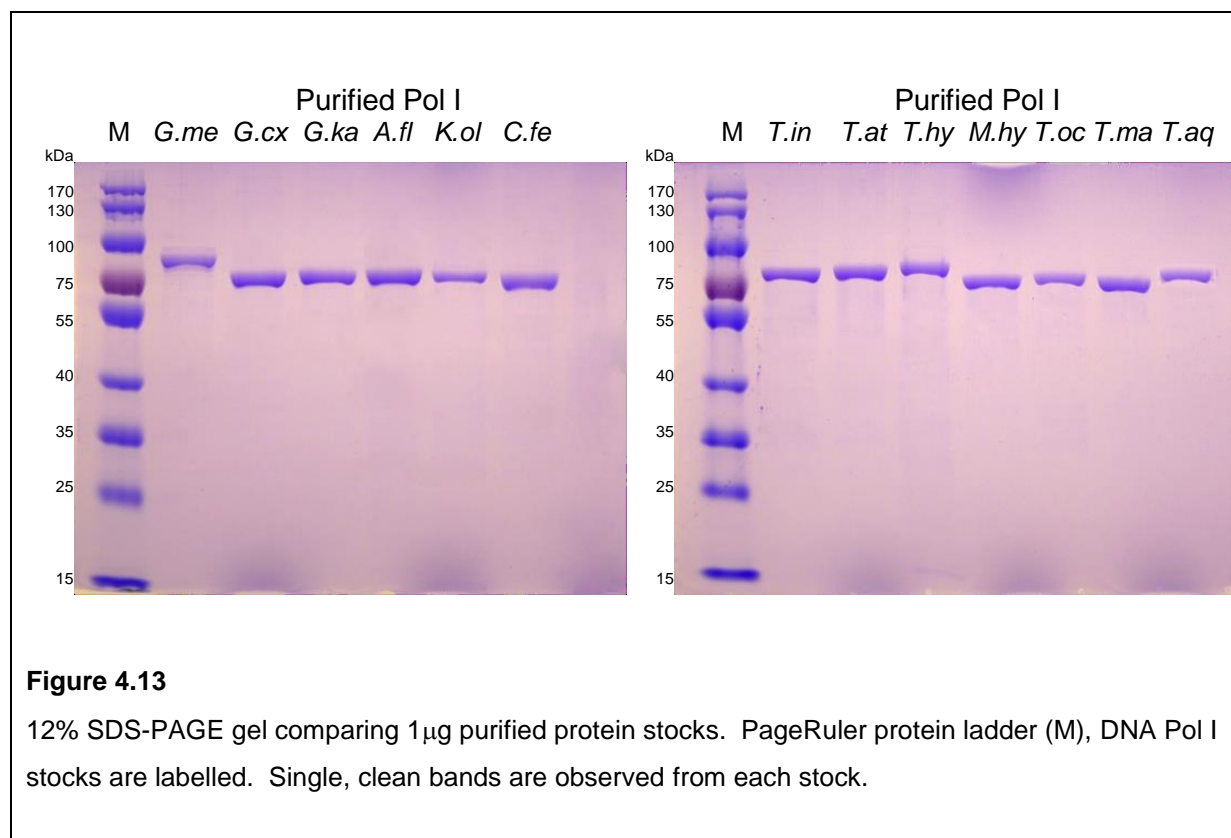
Each DNA Pol I was successfully purified using the three column approach to achieve protein that was shown to be essentially free from background *E.coli* proteins (Figure 4.13). All samples were assumed at this stage to be free from contaminating *E.coli* gDNA although this would need to be confirmed for enzyme characterisation. The pure protein samples were quantified using a fluorescent Qubit® protein concentration assay. BSA standards were supplied at 0, 200, and 400ng/μl by the manufacturer (Invitrogen, UK). 10μl standards/enzyme dilutions were mixed with the Qubit® working solution and incubated at room temperature for 15 min. The plate was read using a Bio-tek fluorimeter, reading at 480nm excitation, 560nm emission. Readings were then converted to mg/ml (μg/μl) using the BSA standard curve to calculate the protein concentration.

DNA Pol I	Protein yield (mg/L)
<i>G.me</i>	35.0
<i>G.ka</i>	42.5
<i>G.cx</i>	42.5
<i>A.fl</i>	11.0
<i>K.ol</i>	33.0
<i>C.fe</i>	30.0
<i>T.in</i>	37.5
<i>T.at</i>	37.5
<i>T.hy</i>	11.0
<i>T.oc</i>	16.0
<i>M.th</i>	40.0
<i>T.aq</i>	33.0
<i>T.ma</i>	20.0

**Table 4.3**

Comparison of DNA Pol I protein yields obtained from 100ml TB culture.

Table 4.3 details total protein yields from the 100ml expression culture (mg/L). Each protein was diluted to 1mg/ml as a working stock for future assays and 1µg purified protein was visualised on a 12% SDS-PAGE gel for comparison (Figure 4.13).



## 4.4 Discussion

The Large Fragment *polA* genes, identified in chapter 3, required a suitable cloning and expression system to facilitate (i) directional cloning, (ii) over-expression of the recombinant protein in a compatible *E.coli* expression strain, and (iii) simple and efficient purification of the over-expressed protein. The pET<sup>®</sup> system was chosen in combination with *E.coli* KRX\_pRARE2, thereby offering high levels of protein expression with tight control of induction.

The use of genetic stop codons in bacteria has been reported to be biased (Vlasov *et al*, 2012). Previous studies at GeneSys Biotech Ltd. have confirmed the advantageous use of TAA as the preferred terminal codon for effective transcription termination. Opal (TGA) and amber (TAG) codons showed an increased occurrence of translational read-through when compared to the ochre (TAA) codon. For this reason, all clones were optimised to introduce an ochre stop codon where not present in the wild-type *polA* sequence.

The Large Fragment Pol I from the twelve thermophilic organisms were successfully cloned and over-expressed. Pol I enzymes were 100% soluble using the expression strain with induction at 24°C. The reduced induction temperature (KRX Induction manual, Promega, UK), in combination with a 6X histidine-tag (Vaugh *et al.*, 2005) were expected to aid solubility of the induced proteins. Two additional *polA* clones were donated to be used as comparison enzymes for this study. The Large Fragment Pol I from *T.aq* and *T.ma* were over-expressed and purified in an identical manner.

A large scale culture induction was used to over-express the Pol I enzymes, providing sufficient protein stocks for downstream characterisation assays. A three column chromatography approach was demonstrated to successfully remove *E.coli* background proteins, purifying each Pol I to apparent homogeneity. A 65°C heat step, prior to purification, was shown to reduce the background *E.coli* proteins, aiding downstream purification.

*Bacillus furnas* (*B.fs*) Pol I was shown to be fully denatured using a 50°C heat step and was therefore not carried forward for purification and downstream characterisation. The *B.fs* strain was originally isolated in 2010 at a temperature of 58°C from thermal run-off water in the Azores, Portugal. The strain was revived in the laboratory on LB agar plates at a maximum temperature of 50°C (reported in Chapter 3), indicating the strain was not as thermophilic as anticipated from the initial sampling temperature. Although lower

temperature enzymes may be suitable for isothermal amplification systems (e.g. Phi29 DNA Pol at 30°C) higher temperatures, at 65°C and above, were desired to provide specific primer annealing and to avoid false priming during reactions. High primer specificity is especially important for diagnostic detection applications, including single-nucleotide polymorphism (SNP) detection where single base discrepancy between sequences may be important.

Metal affinity His-tagged chromatography provided simple and effective first stage purification. The 6x Histidine-tag at the N-terminal domain of each Pol I was not expected to alter the activity of the protein, yet may provide the benefit of increased protein solubility within the *E.coli* expression strain. The Ni-NTA HisPur<sup>TM</sup> column was equilibrated in a purification buffer containing 5mM imidazole. The presence of a low level of imidazole was expected to prevent the weakly interacting *E.coli* proteins from binding the Ni<sup>2+</sup> ions on the column preferentially over the His-tagged Pol I proteins. The use of a low imidazole buffer was especially important when purifying a total protein yield relatively close to the maximum binding capacity of the column. The Ni-NTA HisPur<sup>TM</sup> columns have a reported binding capacity of 60mg/ml (reported for BSA), and were shown to be suitable for the 100ml TB expression cultures used in this project. Flow through fractions did not contain visible Pol I, confirming the binding capacity of the column had not been exceeded.

Heparin column chromatography was chosen as a second stage purification method to take advantage of the nucleic acid binding nature of the DNA Pol I. Purified fractions were shown to be essentially free from background *E.coli* proteins, but a third and final polishing chromatography step was chosen to purify the Pol I to homogeneity, and further concentrate the protein samples. Ion exchange chromatography fractions were shown to contain only Pol I as the visible protein band, confirming a successful purification protocol.

The Qubit<sup>®</sup> protein concentration method was chosen over traditional protein calculation methods; for example, the Lowry assay (Lowry *et al.*, 1951) and Bradford assay (Bradford, 1976). The fluorescent protein method is described as the most sensitive method to calculate the accurate detection of protein well below the limit of UV detection methods and across a wider pH range (Invitrogen, UK). Accurate protein concentrations enabled total yields to be calculated in mg/L culture. High protein yields were achieved for all Pol I enzymes, between 11-42.5mg/L culture, using the T7 expression system. Single protein bands were observed when 1µg protein was electrophoresed on a 12% SDS-PAGE gel.

The purified enzymes reported in this chapter can now be taken forward to investigate their activity in a number of characterisation assays, reported in the next chapter.



## 5. CHARACTERISATION

### 5.1 Introduction

A variety of assays can be applied to characterise and compare the activities of the enzymes in this study. A novel DNA polymerase can be defined as one offering potential advantages over those currently reported in the literature, or one enabling potentially new applications. Comparisons to commercially available polymerases, including *B.st*, *T.aq* and *T.ma* LF DNA Pol I may highlight unique activities. An overview of the techniques used is outlined in the introduction to this chapter.

#### 5.1.1 Polymerase unit concentrations

DNA polymerases, isolated from a wide variety of organisms, vary in their 5'-3' DNA polymerase activity, extension and processivity rates, and with their strand-displacement activity. These activities will vary further with the individual temperature optima and buffer requirements for each enzyme. With this in mind, it was decided to compare each polymerase identified in this study based on a defined quantity of protein. In this way, assays could be readily and easily reproduced. Commercial unit definitions vary greatly, as do the methods for their definition. Assays were therefore performed at a variety of temperatures, with varying protein concentrations, to find the optimal conditions for each enzyme under test. The condition could then be defined as 'optimal' for each enzyme.

#### 5.1.2 Protein thermostability

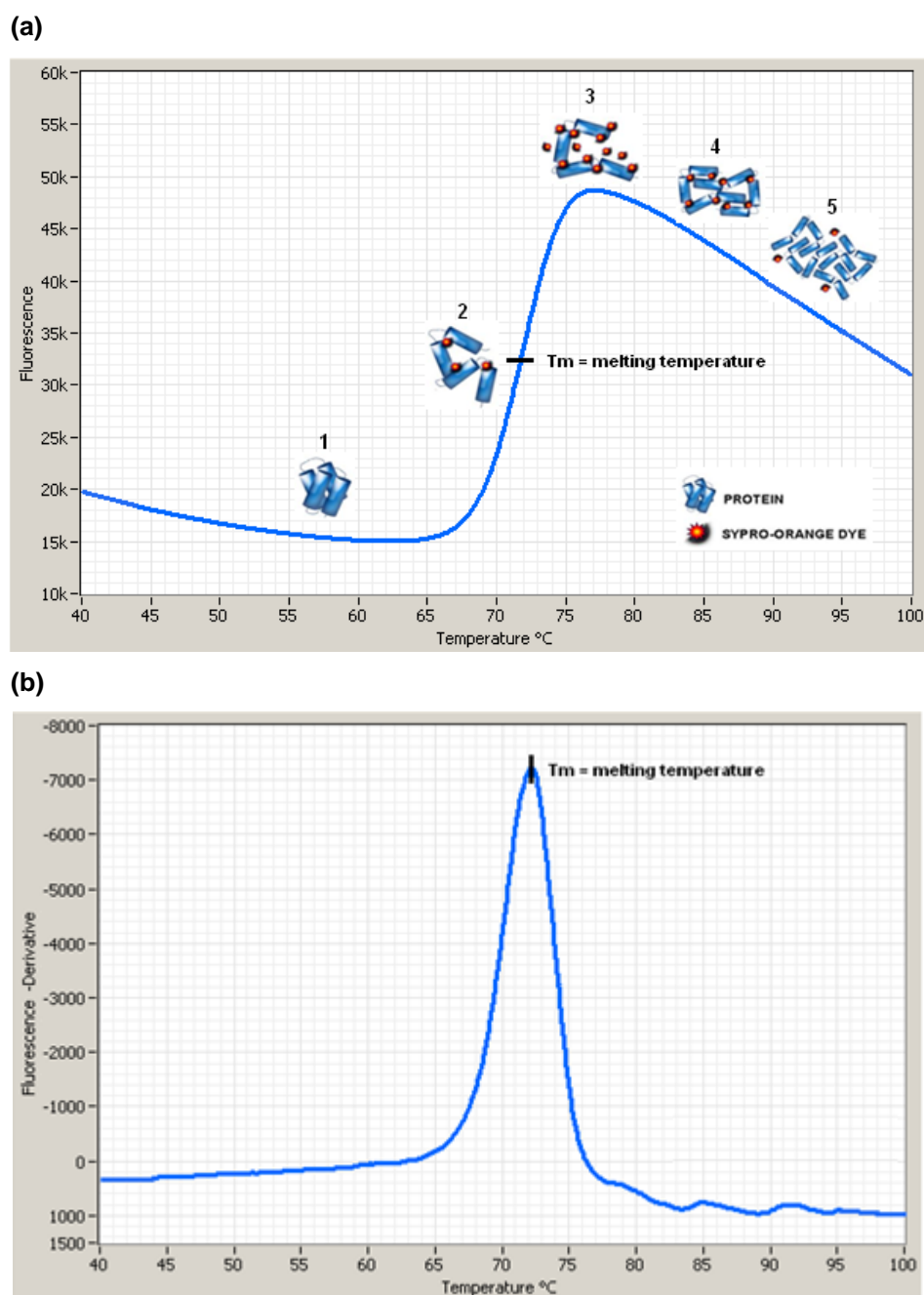
Life can thrive in the most challenging environmental conditions found on planet Earth. The microbial communities that can grow optimally at these extremes of temperature, acidity, salinity, and pressure are classified as extremophiles. This ability requires the stabilisation of all cellular components, so that their functionality is maintained under conditions that would be harmful for most non-extremophile molecules (Danson *et al.*, 2008). Their stability and activity at extreme conditions make them useful for a wide variety of applications in biotechnology.

Life found at high temperature is often associated with extremes of pH, salinity and pressure and are represented in bacteria and archaea. Mesophiles grow optimally between 20-50°C,

'moderate thermophiles' between 50°C and 60°C, 'thermophiles' between 60-80°C, and 'hyperthermophiles' at 80°C to >110°C (Kumar *et al.*, 2011). Life has adapted to these environments enabling their biomolecules to function under these extreme conditions (Hough *et.al.*, 1999).

The general biochemistry of mesophiles and hyperthermophiles have been shown to be very similar. Mesophilic and hyperthermophilic proteins contain the same amino acid building blocks and so it is the interactions between these residues that are the key feature for increased stability. A variety of features have been identified as possible contributors to protein stability at high temperatures and include: (i) increased electrostatic interactions (hydrogen bonds and ion pairs) and disulphide bridges, (ii) a reduced solvent-exposed hydrophobic surface, (iii) reduction in flexible regions (Loops and N- and C-termini regions) through the shortening or anchoring to the protein (Thompson *et al.*, 1999). As a general rule, hyperthermophilic enzymes are more rigid and compact than their mesophilic homologues at mesophilic temperatures and the rigidity is a pre-requisite for high protein thermostability (Danson *et al.*, 1996; Das *et.al.*, 2000).

The thermal denaturation properties of the enzymes in this study can be monitored using a fluorescence based thermal shift assay. The fluorescent dye SYPRO<sup>®</sup>-Orange (Life Technologies, UK) binds specifically to the hydrophobic residues of a protein. Protein denaturation, or 'melting', can be monitored in real-time when a protein sample is incubated with the dye in an instrument capable of sample temperature control and fluorescence detection. With protein unfolding, the hydrophobic residues within the enzyme core are exposed, binding the dye. This results in a significant increase in fluorescence emission (Figure 5.1). Using a derivative plot (rate of change of fluorescence with respect to change in temperature) the melting temperature ( $T_m$ ) can be assigned that is a unique characteristic of the protein in its environment. The thermal profile of each polymerase is expected to reflect the optimal growth conditions of the organism from which it has been isolated, although this will need to be confirmed with each individual polymerase identified.

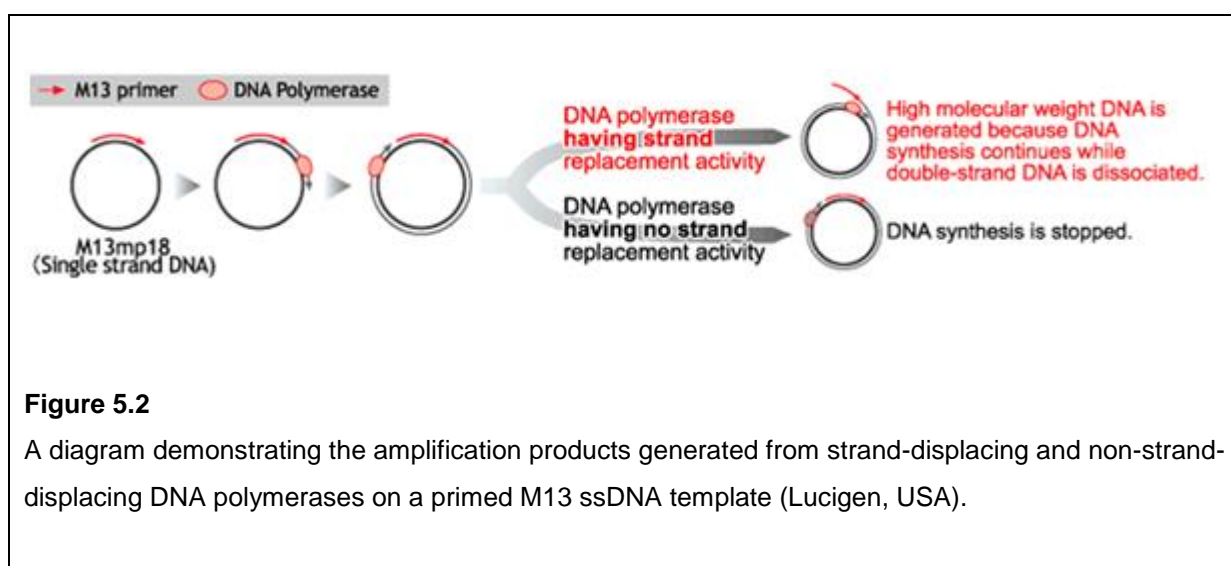


**Figure 5.1**

Overview of the protein thermal shift assay with the intercalation of SYPRO<sup>®</sup>-Orange. (a) Fluorescence measured over an increasing temperature. Protein melting (1), Dye binding (2), Fluorescence peak (3), Protein aggregation (4), Dye dissociation (5). The melting temperature ( $T_m$ ) is assigned by the maximal rate of change of fluorescence. This can be visualised in (b) where a derivative plot (the temperature at which the rate of change in fluorescence with change in temperature is at a maximum) identifies the  $T_m$ .

### 5.1.3 Strand-displacement activity

The focus of this thesis is to identify new DNA polymerases suitable for strand-displacement isothermal amplification reactions. A suitable assay will be required to detect this activity independently of its inherent 5'-3' polymerase activity. These separate activities can be identified using a circular ssDNA template such as the *E. coli* bacteriophage M13mp18 ssDNA. The ability of the polymerase to extend the 3' end of a pre-annealed oligonucleotide primer will confirm an active enzyme, i.e. 5'-3' polymerase activity. The strand-displacement action of a DNA polymerase can then be demonstrated if the primer is further displaced, permitting continued amplification of the template. This end-point primer extension assay will allow each polymerase to be characterised under a variety of conditions, enabling the inherent strand-displacement activity to be visualised through the formation of large amplification products (Figure. 5.2).



**Figure 5.2**

A diagram demonstrating the amplification products generated from strand-displacing and non-strand-displacing DNA polymerases on a primed M13 ssDNA template (Lucigen, USA).

Polymerases showing strand-displacement activity can be further studied using loop-mediated isothermal DNA amplification (LAMP), detailed in Chapter 1.2.3. The previously mentioned M13mp18 method uses a ssDNA template with a primer pre-annealed to it, ready for the polymerase to extend. The LAMP reaction, however, requires the polymerase to initially displace the dsDNA template to provide target primer annealing. Amplification will therefore only occur if sufficient strand-separation activity is present.

LAMP products are characteristically reported by electrophoresing the amplified reaction on an agarose gel that incorporates a DNA-interacting dye, such as ethidium bromide or SYBR® Green I. Unlike PCR, where two primers amplify a sequence of specific size, LAMP

produces multiple copies of the same amplicon, forming large molecular weight structures of varying sizes (demonstrated later in the chapter). As such, a reaction can be said to have amplified, but the specificity of the amplification is difficult to confirm. If a unique restriction endonuclease site is located within the sequence (or deliberately placed within a primer region), a RE digest of the final LAMP product produces a single isolated band, confirming the correct product has been amplified (Notomi *et al.*, 2000).

An alternative detection method monitors the turbidity of the reaction during amplification. Due to the rapid rate of amplification, large amounts of magnesium pyrophosphate by-product are produced, making the reaction cloudy. This enables turbidometric detection to be monitored in real-time (Mori *et al.*, 2001). However, a significant disadvantage of the turbidometric method is the inability to confirm the reaction specificity; the amplification is a yes/no event, and therefore cannot distinguish between false positive results.

Real-time amplification can also be detected when a polymerase is assayed in the presence of a fluorescent dsDNA intercalating dye, such as SYBR<sup>®</sup> Green I or EvaGreen<sup>®</sup>. Much like the intercalation of dyes during a real-time PCR (RT-PCR) reaction, dsDNA intercalating dyes can be used to monitor the amplification of real-time LAMP reactions. A significant advantage of real-time fluorescence detection is the ability to confirm the correct product has been amplified using melt-curve analysis. The exact melting temperature is unique to each amplicon (attributed to its GC content) thus enabling a post amplification confirmation step in RT-PCR (Wittwer *et al.*, 2006). One commercial constraint when using a ‘thermal melt’ profile are the associated license fees on the thermal melt analysis patent. This may deter potential applications from using fluorescent LAMP detection. An alternative analysis method, first identified by OptiGene Ltd., is to use the ‘anneal’ reaction profile of the LAMP amplification product. This profile produces an anneal temperature value ( $T_A$ ) that is highly specific for each target amplicon.

#### **5.1.4 Chapter overview**

This chapter details the results of a variety of enzyme characterisation assays using the purified enzymes from Chapter 4. Real-time fluorescence detection, with anneal curve product confirmation, will be used to characterise the activity of each polymerase in LAMP. The results in this chapter enable direct comparisons of the activities of the enzymes and highlight the key differences between them.

## 5.2 Results

The purified polymerases identified in Chapter 4 were stored as concentrated stocks and required dilution for accurate and reliable comparisons. Each novel polymerase may vary in a wide range of activities. These include 5'-3' DNA polymerase activity, 3'-5' exonuclease activity, and strand-displacement activity. It was therefore decided to compare each enzyme using a defined quantity of protein sample. The gold standard enzyme for isothermal DNA amplification reactions, and the basis for this study, is *B.st* LF Pol I. With a reported specific activity of 120,000 units/mg (NEB, UK), 8u *B.st* Pol I is the equivalent of 66ng protein. The polymerases in this study were therefore compared directly to a '*B.st* Pol I unit equivalent protein' to enable simple and reproducible comparisons throughout.

### 5.2.1 Nuclease contamination

The majority of characterisation assays used in this study require the use of nucleic acid targets. It was therefore necessary to first confirm the absence of nucleases within the purified polymerase samples. 400ng of each polymerase (50u *B.st* Pol I equivalent) were incubated in the presence of a variety of nucleic acid templates. Reactions were performed at 37°C for 3h, in an isothermal reaction buffer (iBuffer, OptiGene Ltd, UK). The reaction was electrophoresed on an agarose gel to compare fragmentation patterns to a no-enzyme control reaction (Figure 5.3).

#### *Exonuclease activity:*

Incubation of each enzyme in the presence of 500ng *Hind* III/*Eco*R I digested lambda DNA resulted in no detectable smearing of bands on an agarose gel (Figure 5.3a).

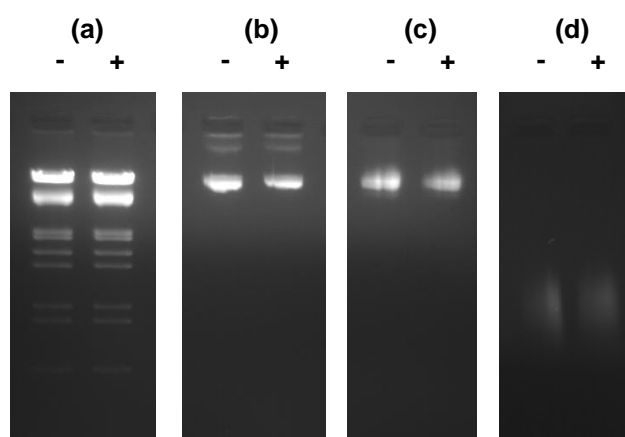
#### *Endonuclease activity:*

Incubation of each enzyme in the presence of 500ng supercoiled pET24a+ dsDNA resulted in no detectable conversion to open-circular or linear forms by agarose gel electrophoresis (Figure 5.3b).

#### *DNase and RNase activity:*

Incubation of each enzyme in the presence of either 500ng M13mp18 ssDNA or MS2 RNA resulted in no detectable degradation or fragmentation of either nucleic acids by agarose gel electrophoresis (Figure 5.3c/d).

Further characterisation assays could now be performed with the knowledge that the polymerase stocks were essentially free of contaminating nucleases.



**Figure 5.3**

Nuclease assay containing 500ng nucleic acid template a/b/c/d, incubated at 37°C for 3h in the presence of 400ng *G.me* Pol I (50u *B.st* Pol I equivalent). (a) 500ng *Hind* III/*EcoR* I digested lambda DNA, (b) 500ng supercoiled pET24a+ dsDNA, (c) 500ng M13mp18 ssDNA, (d) 500ng MS2 RNA.

## 5.2.2 DNA contamination

During the purification of the DNA polymerases there will inevitably be some co-purification of genomic DNA from the *E.coli* strain in which the protein was expressed (Spangler *et. al*, 2009). The ion exchange purification column was expected to have purified significant amounts of enzyme from the nucleic acids, but the inherent nature of a DNA-binding protein infers the protein may bind and carry nucleic acids through the purification procedure. The extent of this contamination needed to be determined to ensure further nucleic acid based characterisation assays were not affected.

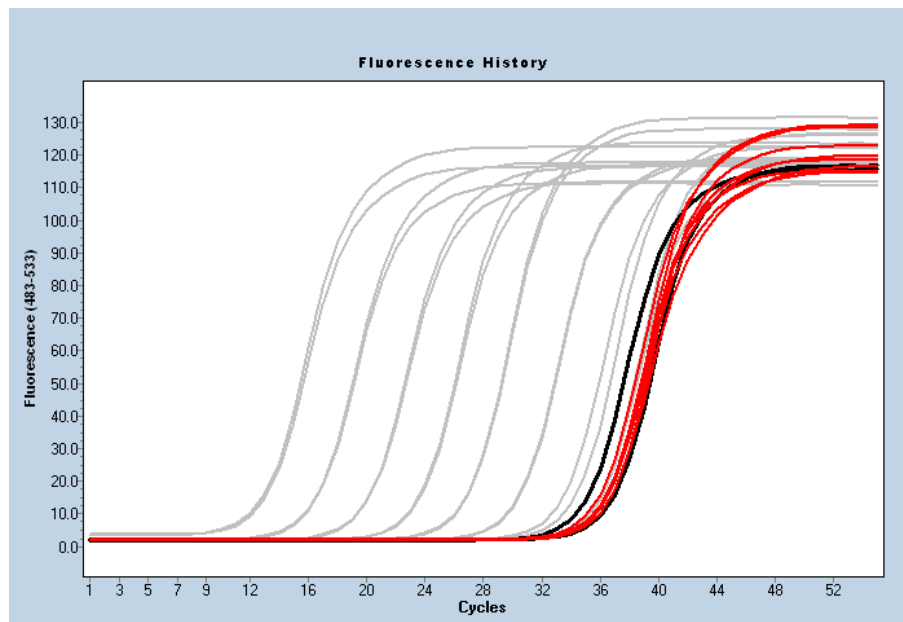
A 23S rRNA quantitative PCR (qPCR) assay was performed to quantitate the level of background *E.coli* gDNA contamination in the purified protein stocks. It was essential to first heat denature the sample at 100°C for 10min, to ensure DNA polymerase activity from the stock itself did not interfere with the qPCR reaction. It was assumed, and later confirmed in this chapter, that the denaturation was sufficient to denature each enzyme. 400ng heat-

treated polymerase (50u *B.st* Pol I equivalent) was added directly to the reaction mix, with a template dilution of *E.coli* gDNA (strain W1485) of known concentration for quantification. The qPCR reaction was performed using a commercial *T.aq* Pol-based mastermix, containing a fluorescent intercalating dye (SYBR<sup>®</sup> Green I) and primers targeting the bacterial 23S rRNA gene (GeneSys Biotech Ltd., UK). Real-time analysis was monitored on a LightCycler<sup>®</sup> 480 qPCR instrument (Roche, USA) following the method described in Chapter 2.6.2. The grey coloured amplification curves seen in Figure 5.4a represent a 1 in 10 dilution series of the *E.coli* gDNA standard (10ng-1fg) with 'no template control' (NTC) reactions highlighted in black. 7 copies of the 23S rRNA genes are present in the *E.coli* genome, quantifying the dilution series from  $1.4 \times 10^7$  to  $1.4 \times 10^{-1}$  copies. The red amplification plots seen in Figure 5.4a represent the level of background *E.coli* gDNA contamination in 400ng of the protein samples. If carryover *E.coli* gDNA was present, amplification would occur earlier than that of the level of the NTC. This is not observed. All Pol I samples show amplification at the level of  $1.4 \times 10^{-1}$  copies, i.e. the NTC in the commercial *T.aq* Pol-based mastermix. This result shows there is no additional background *E.coli* gDNA present in the protein samples.

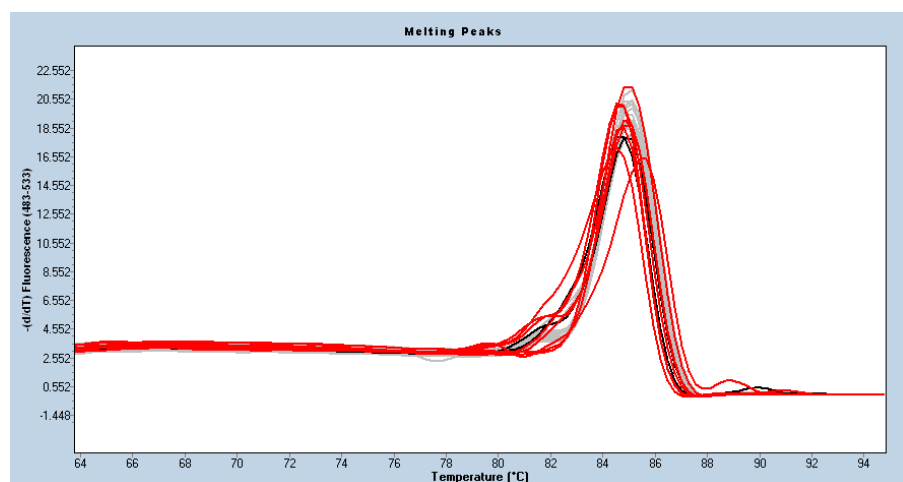
To confirm the specificity of the 23S rRNA amplification, a melt curve was performed on each qPCR amplicon (Figure 5.4b). All amplicons showed a melt curve with a peak of 84.6°C (+/- 0.1°C), previously confirmed to be specific for the 23S rRNA gene (GeneSys Biotech Ltd., UK).



(a)



(b)



**Figure 5.4**

qPCR *E. coli* gDNA contamination assay. (a) qPCR amplification of the 23S rRNA target using an *E. coli* gDNA template dilution ( $1.4 \times 10^{7,6,5,4,3,2,1,0,-1}$  copies) run in duplicates (grey), NTC (black), heat-treated protein samples (400ng per reaction) (red). All samples amplify at the  $1.4 \times 10^{-1}$  copy dilution, with the NTC. (b) Melt-curve analysis of the amplified product, coloured as above, shown to all melt at the same temperature, indicating identical amplification products have been amplified.

### 5.2.3 Protein thermal melt

The qPCR results confirmed the purified stocks were gDNA contaminant free, and were therefore not expected to have a significant effect on further nucleic acid based characterisation assays.

A thermostability profile of each polymerase was required to provide information on the maximum temperature for further activity characterisation assays. The thermal profile of an enzyme is dictated by its reaction environment, including interactions with additional proteins if present, and therefore the choice of reaction buffer used was crucial to obtain meaningful results. All commercially available DNA polymerases are supplied with a reaction buffer suitable for its specific reaction activity. A comparison by GeneSys Biotech Ltd. of available thermostable polymerase buffers, required for nucleic acid amplification reactions, identified common components. These include Tris-HCl pH7.5-9.0 (@ 25°C), monovalent ions, including 10-60mM KCl, 10-60mM (NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> and 3-5mM MgSO<sub>4</sub>. The 10X iBuffer (OptiGene Ltd., UK) was identified as an optimised reaction buffer for isothermal amplification enzymes, with reaction components falling within the common polymerase buffer ranges. The iBuffer was therefore expected to provide suitable conditions for each of the 13 polymerases identified, and was chosen for the thermal melt assay, and for all subsequent nucleic acid amplification assays to ensure consistency.

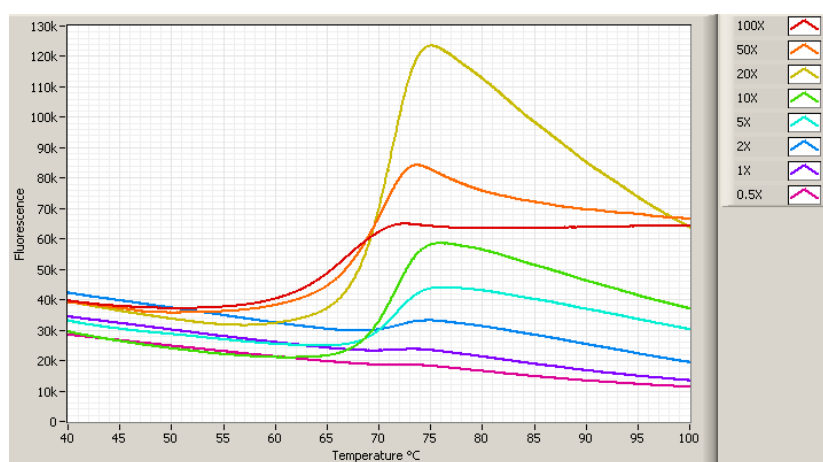
The protein-melt assay required additional development before the thermostability of the purified enzymes could be assessed.

#### Dye concentration optimisation

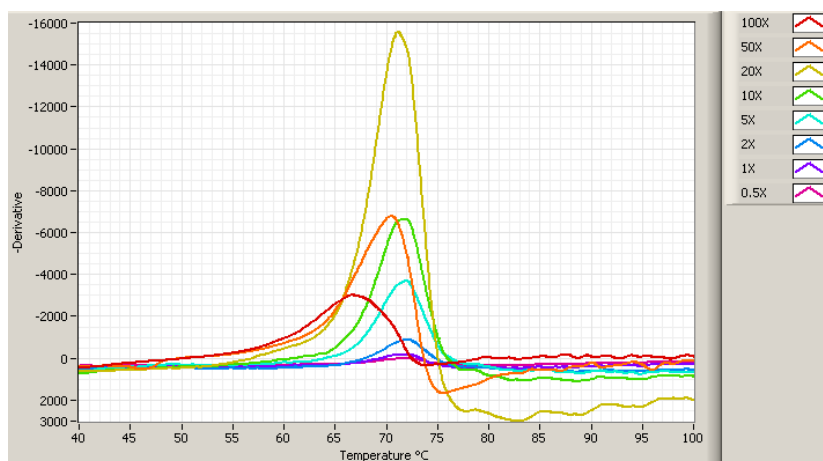
The ratio of SYPRO<sup>®</sup>-Orange to protein sample was investigated to initially confirm the reliability of the assay. A 25µl reaction containing 1µg *G.me* LF Pol I, 0.5-100X SYPRO<sup>®</sup>-Orange and 1X iBuffer was incubated on a Genie<sup>®</sup>II instrument (OptiGene Ltd., UK). The Genie<sup>®</sup>II reports fluorescence at 480nm excitation and 560nm emission, with ramping at 0.05°C/sec from 35°C to 105°C. Fluorescence was reported continuously to provide a complete thermal melt profile for each reaction. Figure 5.5a shows the thermal melt profile of *G.me* LF Pol I, with the data reported in Figure 5.5b as a derivative plot (maximal rate of change of fluorescence over temperature) to identify the peak value (T<sub>m</sub>). Table 5.1 shows the T<sub>m</sub> values obtained for each reaction, comparing the effect of the protein intercalating dye.

The level of SYPRO<sup>®</sup>-Orange used was shown to change the value of the melting temperature of the protein, reporting 71.74°C to 66.65°C, with higher dye concentrations lowering the reported temperature of protein denaturation. Less than 1°C variation was observed between 1-20X dye, and this consistency was taken as an indicator for accuracy over that concentration of the dye. It was noted the Genie<sup>®</sup>II instrument detection limit was  $\pm 0.1^\circ\text{C}$  and may have accounted for the subtle variations recorded. A 10X SYPRO<sup>®</sup>-Orange dilution, the equivalent to a 1/500 dilution of the supplied stock, was chosen for all further assays.

(a)



(b)



**Figure 5.5**

Thermal melt profile with  $1\mu\text{g}$  *G.me* LF Pol I in the presence of a dilution of SYPRO<sup>®</sup>-Orange intercalating dye (0.5-100X final). (a) Real-time thermal melt profile. (b) The real-time thermal melt profile reported as a derivative plot, from which the  $T_m$  values, shown in Table 5.1, were determined.

SYPRO <sup>®</sup> -Orange (final)	Peak Value (T <sub>m</sub> ) (°C)
100X	66.65
50X	70.47
20X	71.14
10X	71.90
5X	71.95
2X	72.07
1X	72.04
0.5X	no data

**Table 5.1**

T<sub>m</sub> values reported for the thermal melt profiles reported in Figure 5.5b

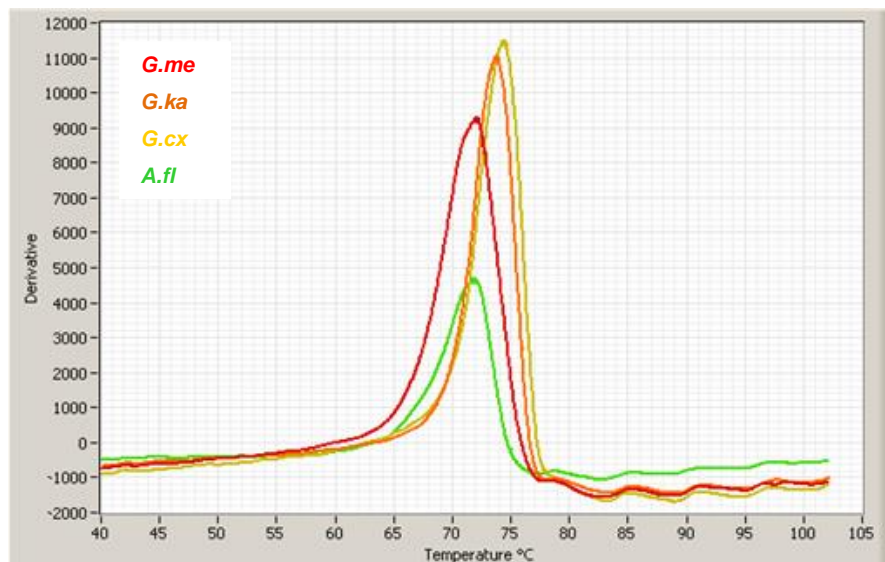
### Polymerase thermostability

1µg protein sample was incubated in a 25µl reaction containing 10X SYPRO<sup>®</sup>-Orange and 1X iBuffer, and analysed on the Genie<sup>®</sup>II as before. Each Pol I sample was shown to melt in the presence of the dye (Figure 5.6). The peak height of the derivative plot was used to define the T<sub>m</sub> for each Pol I for comparison. Using a defined concentration of each protein in the assay, conclusions can be drawn with respect to the width and peak height of each melt profile, assumed to be a characteristic of each polymerase. The width of the peak indicates the temperature range over which unfolding takes place, reflecting the co-operativity of the unfolding process. The more co-operative the process, the faster the denaturation exposes the hydrophobic protein core to the SYPRO<sup>®</sup>-Orange intercalating dye, resulting in a rapid increase in the reported fluorescence, and a narrower peak. Assuming each enzyme of equivalent size will bind approximately the same amount of dye when fully unfolded, the area under the peaks will be approximately the same. Thus, the wider the peak and lower the height of the peak, the slower the denaturation.

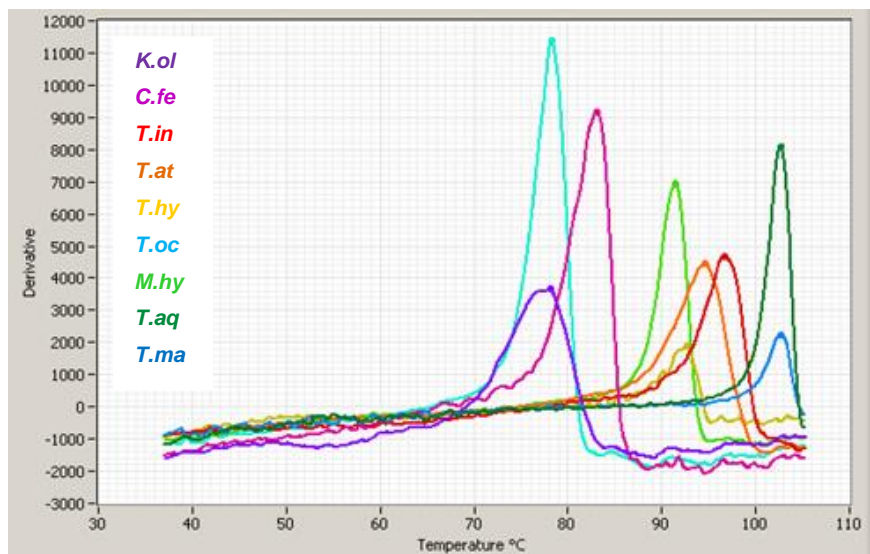
The *Geobacillus*-like (*G.me*, *G.ka*, *G.cx* and *A.fl*) LF Pol I were shown to melt between 64-78°C (Figure 5.6a). *K.ol*, *T.oc*, *C.fe*, *M.hy*, and *T.hy* LF Pol I were shown to be more thermostable, denaturing between 70-95°C. *T.in* and *T.at* LF Pol I were observed to melt between 86-102°C, with *T.aq* and *T.ma* LF Pol I showing the highest thermostability, melting between 95-105°C (Figure 5.6b). All polymerases were shown to be more thermostable than the optimal growth temperature of the associated organism (Figure 5.7). This was not a surprising result with the upper growth limit for all organisms reported to be several degrees (°C) above the optimal growth temperature (Table 3.2). DNA polymerases are an essential

component of an organism's replicative machinery, and therefore its encoded polymerases must be thermostable to at least the upper limit of the organisms growth temperature.

(a)

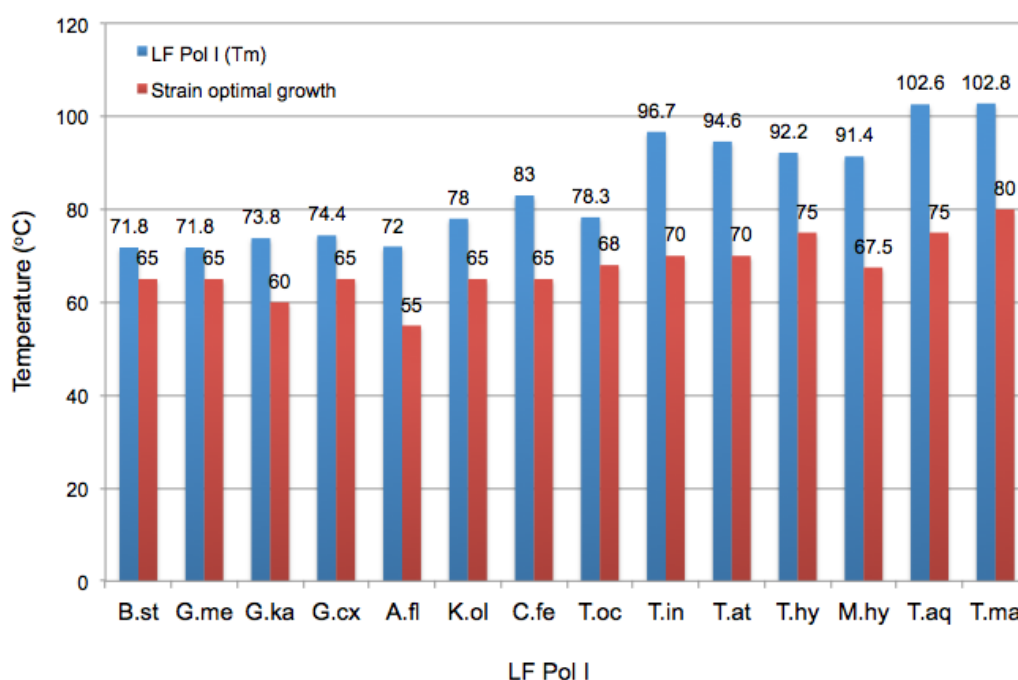


(b)



**Figure 5.6**

Thermal melt analysis of the LF Pol I (a) 1 $\mu$ g thermal melt profile derivative plot of: *G.me* (red), *G.ka* (orange), *G.cx* (yellow), *A.fl* (green), (b) 1 $\mu$ g thermal melt profile derivative plot of: (b) *K.ol* (purple), *C.fe* (pink), *T.in* (red), *T.at* (orange), *T.hy* (yellow), *T.oc* (light blue), *M.hy* (light green), *T.aq* (dark green), *T.ma* (dark blue). Peak dots indicate the  $T_M$  value reported.



**Figure 5.7**

A comparison of the organisms reported optimal growth temperature, as reported in Table 3.1 in Chapter 3 (red bars). The LF Pol I T<sub>m</sub>, reported from the thermal melt analysis in Figure 5.6 (blue bar). The data are reported above individual bars for clarity.

## 5.2.4 Strand-displacement activity

Having identified the temperature profile for each Pol I, further assays could be performed in the knowledge of each upper limit of thermostability. Nucleic acid amplification assays further define the upper and lower reaction limits for characterisation. Too low a reaction temperature may lead to non-specific amplification, including the formation of primer dimers. Too high a reaction temperature will restrict efficient primer/template annealing, potentially preventing amplification entirely. Primer design was therefore critical, with high annealing temperatures required to cover all the reaction conditions to be investigated, up to a maximum of 80°C, to prevent denaturation of the amplicon itself.

### M13mp18 ssDNA primer extension assay

The activity of the Pol I enzymes had not yet been confirmed. An M13mp18 ssDNA primer extension assay was chosen to (i) initially confirm the 5'-3' DNA polymerase activity of the enzyme, and (ii) investigate the ability of each polymerase to strand displace DNA.

Sequence alignments in Chapter 3 identified several of the DNA polymerases to potentially show 3'-5' exonuclease activity. This activity may degrade the single-stranded primers essential to the reaction. The extension primer used in this assay was therefore modified to include a 3'-phosphorothioate base at the n-1 position (Figure 5.8).

**-47\_extension**

5' -GCCGTTGCTACCCCTCGTTCCGATGCTGT<sup>S</sup>T-3'

**5kb\_blocking**

5' -GCCGTTGCTACCCCTCGTTCCGATGCTGTCT<sup>PHO</sup>-3'

**Figure 5.8**

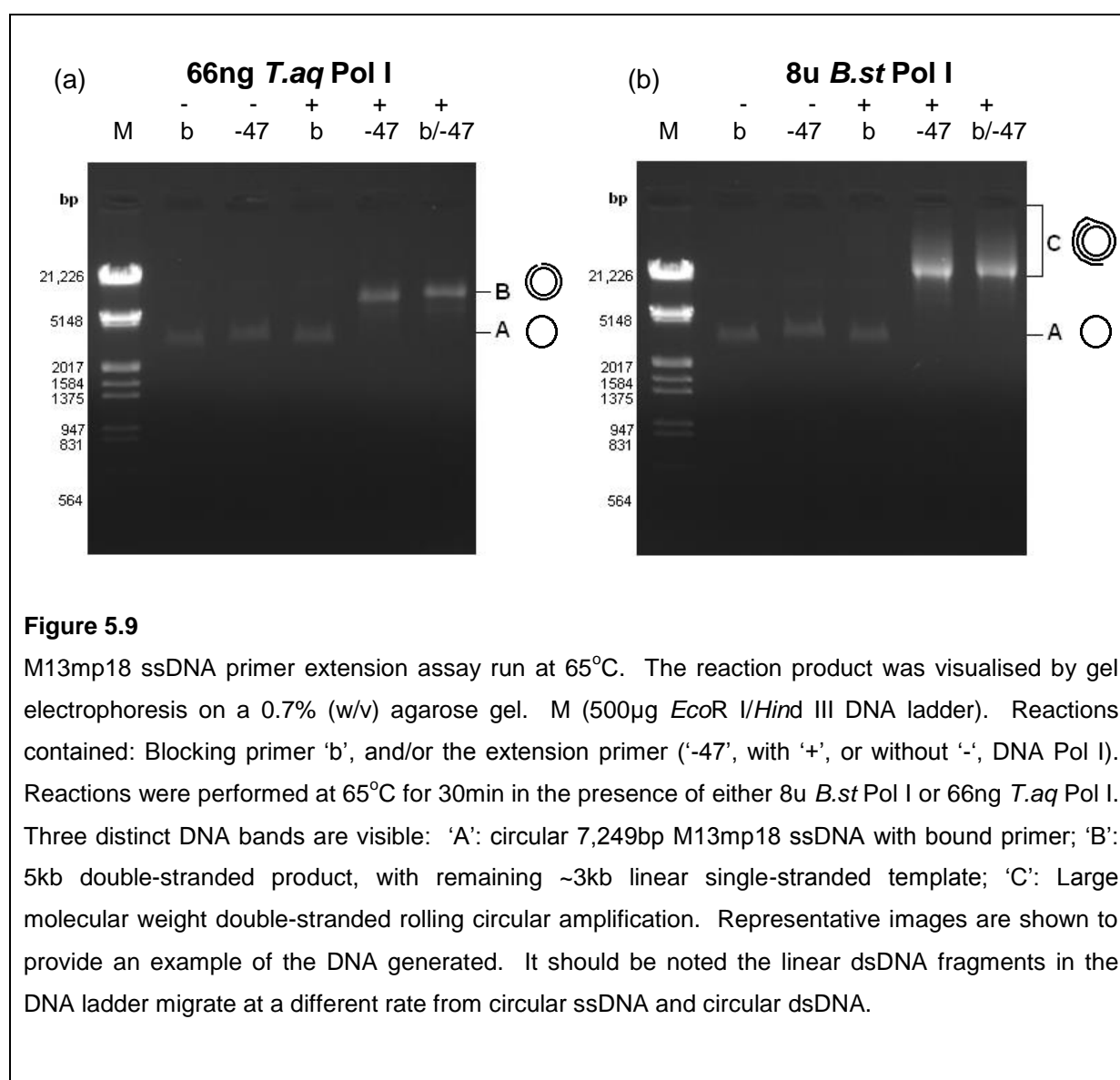
The oligonucleotide primers used in the M13mp18 ssDNA primer extension assay. Primers were designed using Oligo7 (Molecular Biology Insights Software) with annealing temperatures accepted above 80°C. The -47\_extension primer, containing a 3'-end (n-1) phosphorothioate modification (C<sup>S</sup>), enables 5'-3' DNA polymerase primer extension yet prevents 3'-5' exonuclease degradation. The 5kb\_blocking primer, containing a 3'-phosphate group (PHO), prevents primer extension.

The extension assay requires a primer to be pre-annealed to the ssDNA circular template, enabling a DNA polymerase to extend around, generating dsDNA. The method was modified to include an additional pre-annealed primer, with a 3'-end phosphate modification. A 3'-OH group is necessary for primer extension, and therefore a modified 3'-end prevents this activity. This primer therefore acts as a 'blocking' primer, stalling a polymerase that does not have the required activity to displace it. The modification has also been shown to prevent 3'-5' exonuclease degradation (Lehmann *et al.*, 1964). The blocking primer was designed 5kb downstream of the -47\_extension primer (Figure 5.8). Extension of a 5kb product, confirmed by agarose gel electrophoresis, would identify an enzyme showing 5'-3' DNA polymerase activity, but no strand-displacement activity. An enzyme with strand-displacement activity will displace the blocking primer, continuing to amplify around the circular ssDNA template, further displacing the -47 extension primer, and the now dsDNA ahead of it. This rolling circle amplification (RCA) can be confirmed by the generation of a high molecular weight product on an agarose gel stained with ethidium bromide.

Assays were performed at 65°C, 70°C, 75°C and 80°C to compare individual Pol I activities. Primer sequences can be found in Appendix i.vii.

## Proof of principle

It was necessary to test the assay to ensure primers annealed correctly, and that the phosphate modification on the blocking primer prevented extension by the Pol I. It was important to use a molar excess of primers to template in the reaction to ensure each ssDNA strand contained an annealed primer. Control enzymes of known strand-displacement activity were compared, to identify the different amplification fragments. 66ng enzyme (8u *B.st* LF Pol I equivalent) were incubated in a reaction mix, including M13mp18 ssDNA template, pre-annealed with the -47\_extension and 5kb\_blocking primers at 65°C for 30min. Assays were placed on ice and the reaction stopped by the addition of a gel loading solution containing 10mM (final) EDTA, to chelate out  $Mg^{2+}$  from the reaction. DNA products were visualised on a 0.7% (w/v) agarose gel for comparison (Figure 5.9a/b).





The proof of principle identified a variety of single-stranded and double-stranded DNA amplification products. The reaction confirmed:

- Single-stranded DNA does not resolve at the same rate as a double-stranded DNA equivalent. The 7,249bp M13mp18 ssDNA template, with annealed primer, was observed to migrate significantly faster through the 0.7% (w/v) agarose gel.
- When compared to a dsDNA ladder, the intercalating dye (ethidium bromide) was also shown to bind inefficiently to ssDNA, compared to dsDNA, showing brighter fluorescence with dsDNA.
- 5'-3' DNA polymerase activity of the enzyme was confirmed by the generation of dsDNA products, either as a 5kb linear band, or as high molecular weight DNA.
- The phosphorylated blocking primer prevented 5'-3' DNA polymerase primer extension.
- *B.st* LF DNA Pol I demonstrated strand-displacement activity, displacing the blocking primer, to show rolling-circle-like amplification. This was observed by the amplification of high molecular weight dsDNA back up to the well.
- *T.aq* LF Pol I does not show strand-displacement activity, stalling at the blocking primer, to produce a 5kb circular dsDNA product.

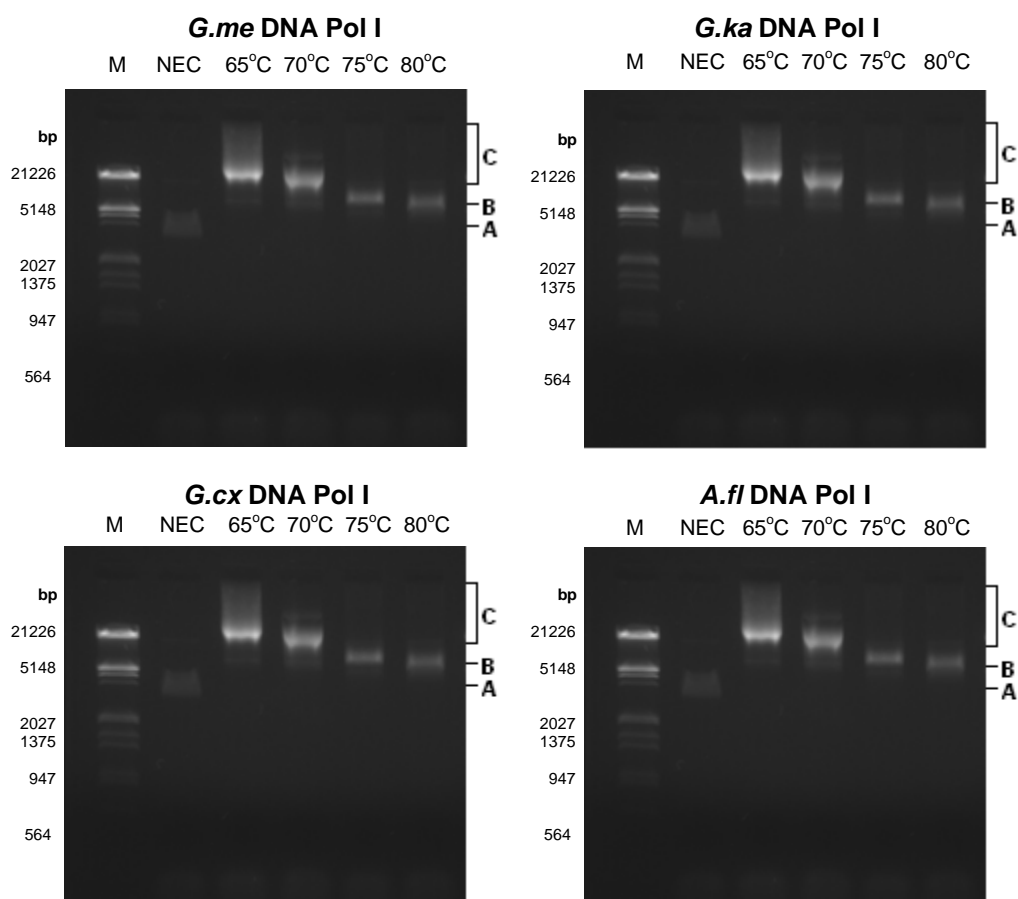
#### Polymerase activity:

The primer extension assay (with blocking primer) was performed with 66ng (8u *B.st* LF Pol I equivalent) of each polymerase at 65°C, 70°C, 75°C and 80° for 30min, on a Veriti® thermal cycler (ABI, UK) to compare the effect of temperature on enzyme activity.

*G.me*, *G.ka*, *G.cx* and *A.fl* Pol I showed strand-displacement activity at 65°C (Figure 5.10). SD activity was also observed at 70°C but less dsDNA product was generated, indicating reduced enzyme activity at the higher temperature. SD activity could not be seen at the 75°C and 80°C, as shown by a dsDNA band from the reaction stalling at the blocking primer. The thermal shift assay showed the *Geobacillus* polymerases to denature rapidly at 78°C, explaining the minimal primer extension observed in the 75°C and 80°C reactions.

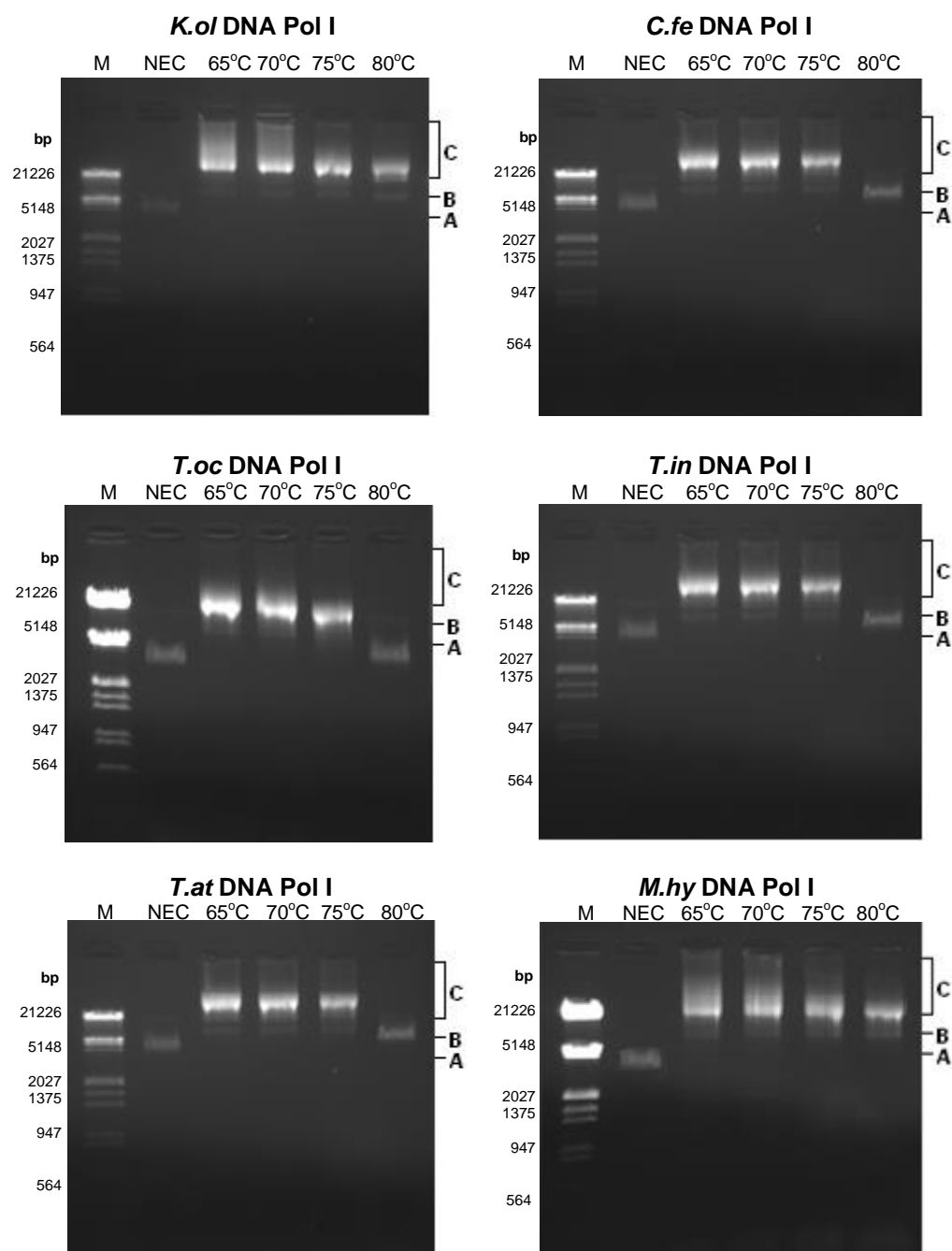
*K.ol*, *C.fe*, *T.oc*, *T.in*, *T.at* and *M.hy* polymerases were further demonstrated to strand displace DNA, with large amplification products observed (Figure 5.11). All polymerases showed maximal SD activity at 65°C, but significant SD activity remained at the higher

temperatures, corresponding to the increased thermostability of these enzymes. *K.ol* and *M.hy* polymerases uniquely showed SD activity at 80°C.



**Figure 5.10**

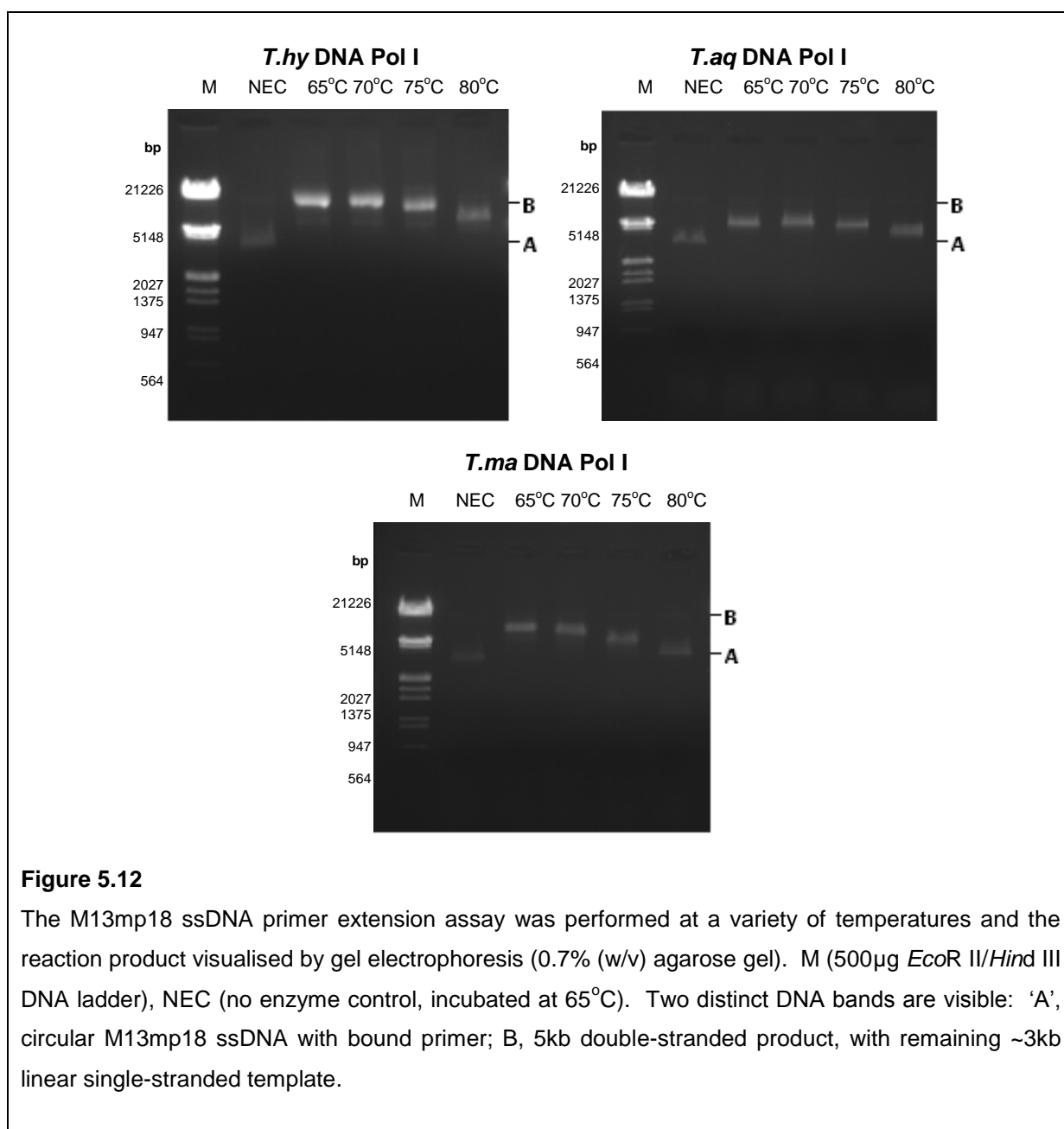
The M13mp18 ssDNA primer extension assay was performed at a variety of temperatures for 30min and the reaction products were visualised by gel electrophoresis (0.7% (w/v) agarose gel). M (500µg *EcoR* I/*Hind* III DNA ladder), NEC (no enzyme control, 65°C). Three distinct DNA bands are visible: 'A', circular M13mp18 ssDNA with bound primer; B, 5kb double-stranded product, with remaining ~3kb linear single-stranded template; C, Large molecular weight double-stranded rolling circular amplification.



**Figure 5.11**

The M13mp18 ssDNA primer extension assay was performed at a variety of temperatures and the reaction product visualised by gel electrophoresis (0.7% (w/v) agarose gel). M (500µg *EcoR* I/*Hind* III DNA ladder), NEC (no enzyme control, 65°C). Three distinct DNA bands are visible: 'A', circular M13mp18 ssDNA with bound primer; B, 5kb double-stranded product, with remaining ~3kb linear single-stranded template; C, Large molecular weight double-stranded rolling circular amplification.

*T.hy*, *T.aq* and *T.ma* Pol I did not demonstrate strand-displacement activity in this assay (Figure 5.12). The enzymes showed 5'-3' DNA polymerase activity, generating a 5kb dsDNA product, indicating successful primer extension, but stalled at the blocking primer. *Thy*, *Taq* and *Tma* polymerases showed reduced primer extension activity with increasing temperature, demonstrating maximal activity at 65°C.



The identification of several thermostable Pol I enzymes, all demonstrating maximal strand-displacement activity at 65°C, indicated their potential use as an alternative to *B.st* Pol I in isothermal amplification reactions. Their suitability for use in a LAMP reaction was therefore investigated.

### 5.2.5 Isothermal DNA amplification

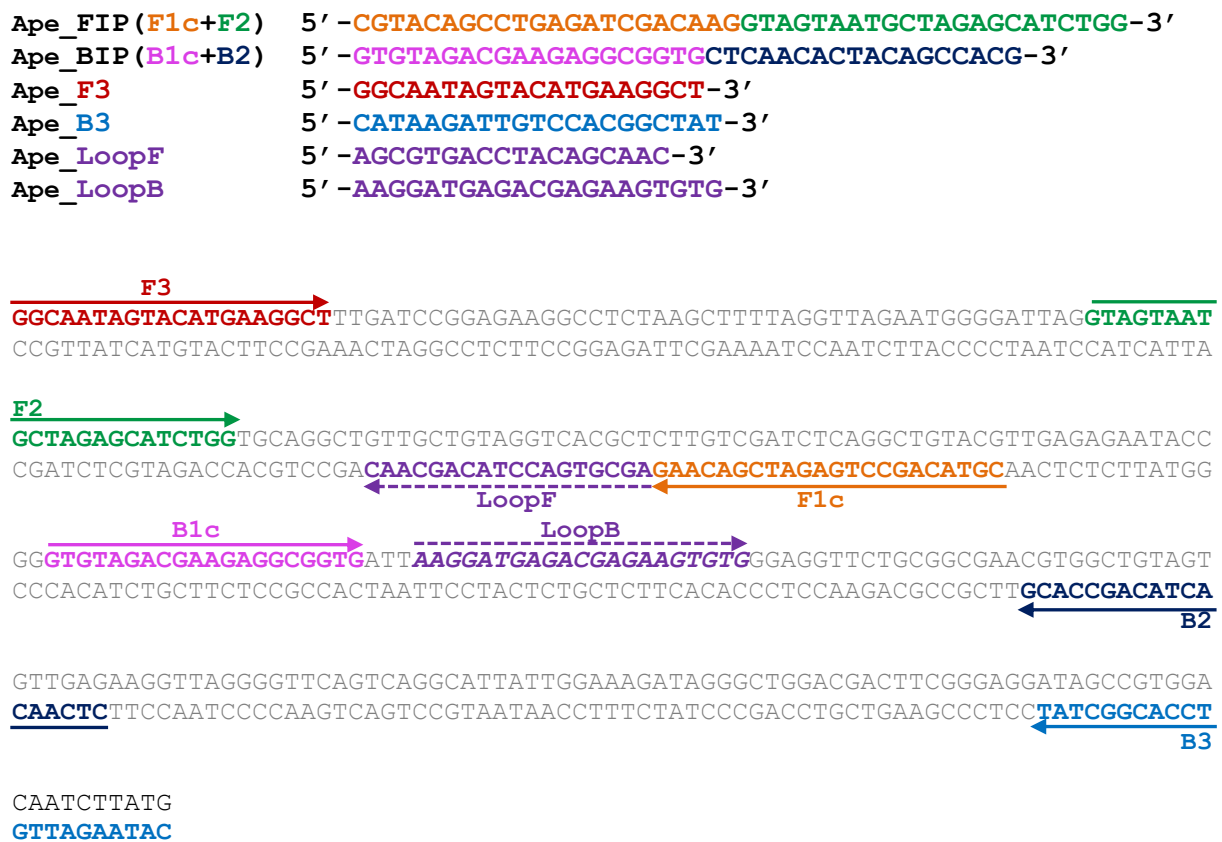
In 2000, Notomi *et al.* published a novel DNA amplification method that was shown to rapidly amplify DNA with high specificity and efficiency under isothermal conditions. Of all the isothermal methods described, LAMP was chosen for this study due to its high publication number, and its reported superior ease of use, reliability and amplification speed.

#### Proof of principle

Before characterisation assays were performed, the LAMP method itself was investigated to gain a greater understanding of this relatively new amplification technique.

#### *Target Design:*

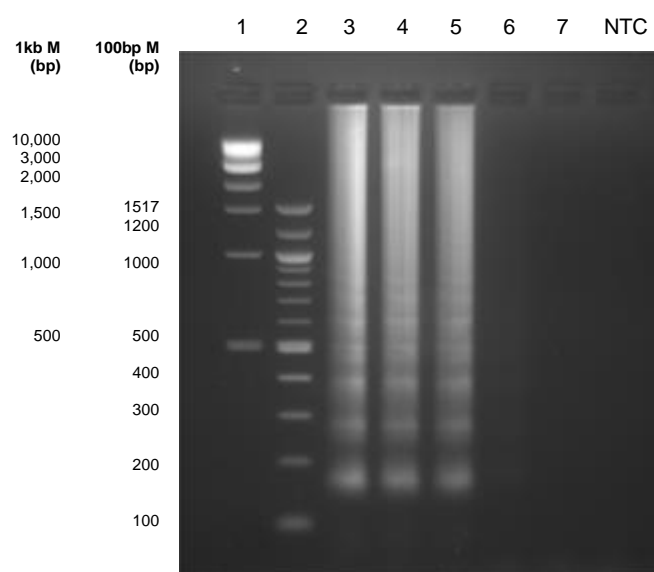
A non-*E.coli* based target was chosen as template to test the LAMP system in order to avoid potential amplification from trace *E.coli* gDNA, if still present within the purified enzyme samples. The *Aeropyrum pernix* (*A.pe*) genome (NCBI\_BA000002) was chosen and subjected to analysis by LAMP primer design software (LAMPdesigner – PremierBiosoft, USA) with design parameters set to yield LAMP primers with an annealing temperature ( $T_A$ ) of 65°C. 6 primers were designed (FIP, BIP, F3, B3, LoopF and LoopB) to amplify a 310bp fragment of *A.pe* gDNA (Figure 5.13). A BLAST<sub>n</sub> search of the DNA fragment showed little identity to any other sequence in the database and was therefore specific to the *A.pe* gDNA template to be used.



**Figure 5.13**

The 310bp *A. pe* target DNA sequence identified for LAMP primer design. The location of the 6 LAMP primers, with 8 specificities, are highlighted on the DNA fragment.

LAMP reactions were run as reported by Notomi *et al.* (2000), detailed in Chapter 2.6.5. A 25µl reaction contained 1X ThermoPol Buffer (NEB), 8u *B.st* Pol I (NEB), 0.4mM dNTP mix (Promega, UK), 1M Betaine (Sigma, UK), 5mM MgSO<sub>4</sub> (Sigma, UK), *A. pe* gDNA (donated by Dr Clark, GeneSys Biotech Ltd, UK) and the *A. pe* LAMP primer mix. Reactions were heated in a waterbath at 65°C for 60 min to allow amplification to occur. Once completed, reactions were heated to 85°C for 10 min to stop the reaction. An aliquot (1/5<sup>th</sup> reaction) was visualised on a 2% (w/v) agarose gel stained with ethidium bromide to confirm amplification (Figure 5.14). The reaction tubes were opened in a separate laboratory to avoid contamination by the amplicon.



**Figure 5.14**

5µl of 25µl reaction were electrophoresed on a 2% (w/v) agarose gel stained with ethidium bromide. Lane 1: 500ng NEB 1kb ladder, lane 2: 500ng NEB 100bp ladder, lane 3-7: LAMP reaction product  $1 \times 10^{5,4,3,2,1}$  copies *A.pe* gDNA template, lane 8: no template control (NTC).

The characteristic LAMP banding pattern observed in Figure 5.14, formed of repeating copies of the same amplicon, indicated successful amplification, with sensitivity achieved down to 100 copies of the *A.pe* gDNA template.

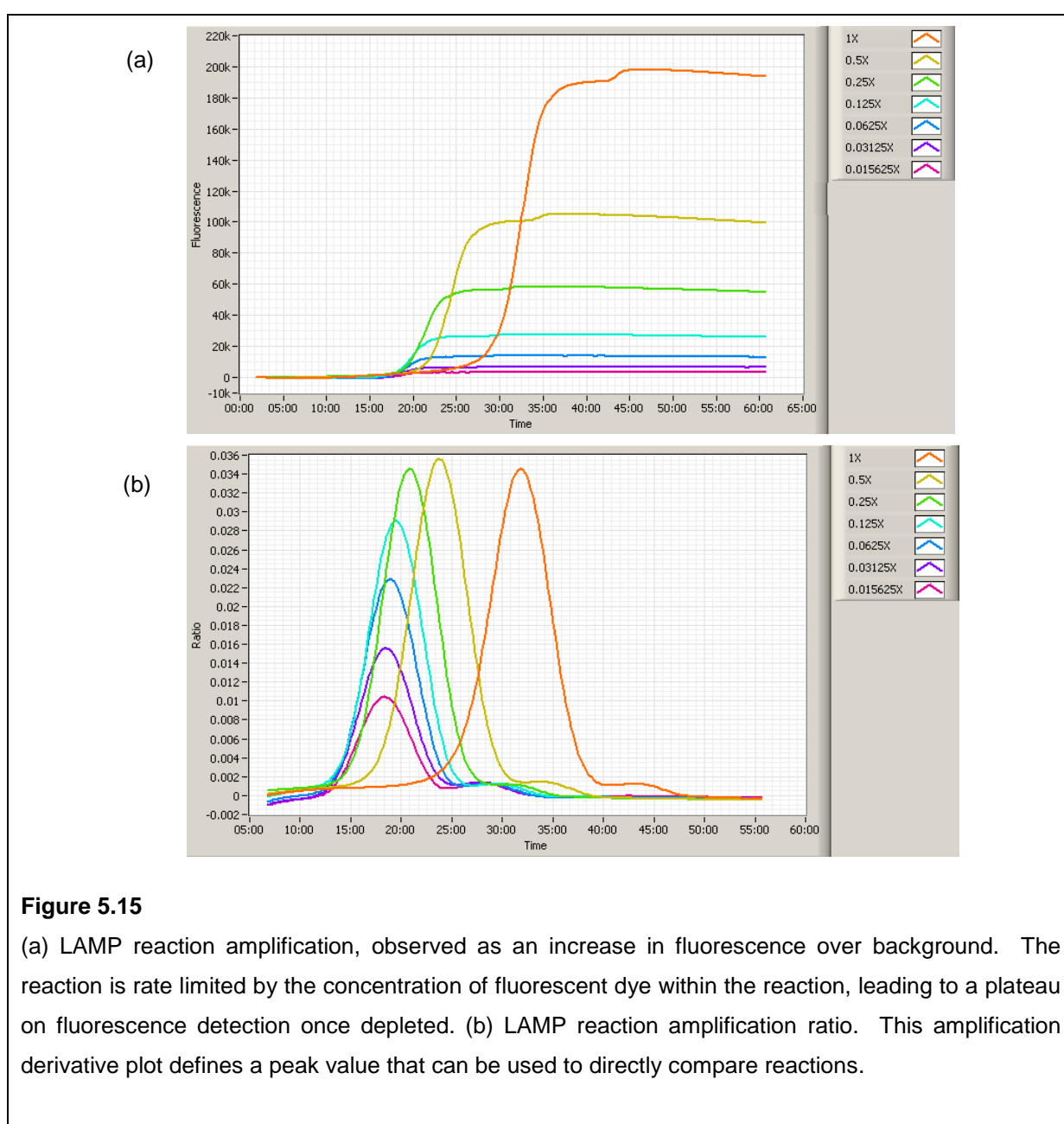
#### *Detection:*

Intercalating dyes are essential for the real-time monitoring of amplification products in a qPCR reaction. Similarly, the same dyes have been successfully used for real-time LAMP detection. The inhibitory effect of the double-stranded intercalating dye Eva Green® (Biotium, USA) was assessed using *B.st* LF Pol I.  $1 \times 10^5$  copies of *A.pe* gDNA were amplified as before but in the presence of a dilution of the dye. Enzyme inhibition, if observed, was expected to result in a longer time to amplification due to a reduced ability to interact with DNA bound to the dye. The fastest time to result therefore determined the optimum concentration of dye to be used. Reactions were analysed on the Genie® II (OptiGene Ltd. UK), monitoring fluorescence every 5 sec at 65°C.

Rapid amplification of the specific target by the polymerase leads to a significant increase in dsDNA. The increased intercalation of the dsDNA specific dye leads to an increase in fluorescence, as observed in Figure 5.15a. The amplification ratio (rate of fluorescence

change over time) can be viewed in Figure 5.15b, enabling a defined amplification time to be reported to the reaction (Table 5.3).

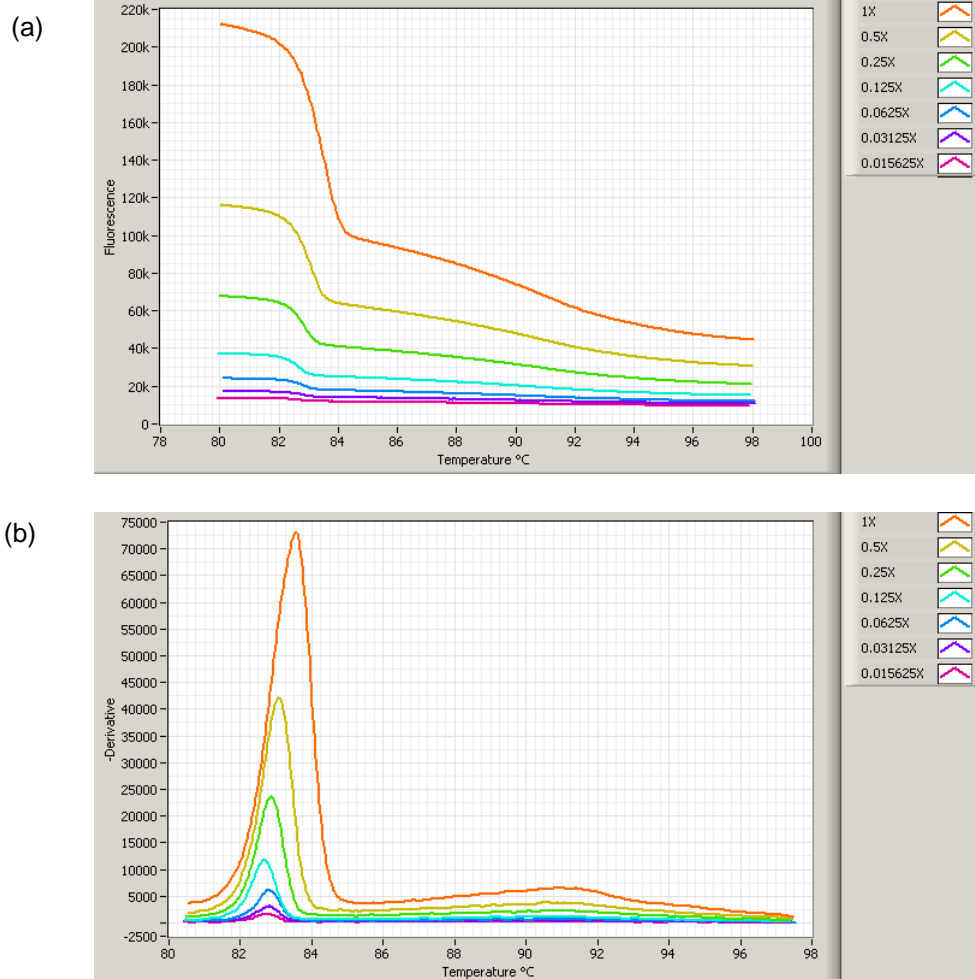
The optimum non-inhibitory concentration of Eva Green® was determined to be 0.125X (final), and was used for all further real-time LAMP reactions. Although this level of dye did not give the fastest possible amplification time (Table 5.3), it was shown to give a comparatively high fluorescence value on the Genie® II (Figure 5.15b). The *A.pe* LAMP amplicon was found to anneal specifically at 82.7°C under the defined assay conditions (Figure 5.16a/b and Table 5.3). This target confirmation step offers a significant advantage over gel-based, end point detection.



**Figure 5.15**

(a) LAMP reaction amplification, observed as an increase in fluorescence over background. The reaction is rate limited by the concentration of fluorescent dye within the reaction, leading to a plateau on fluorescence detection once depleted. (b) LAMP reaction amplification ratio. This amplification derivative plot defines a peak value that can be used to directly compare reactions.





**Figure 5.16**

(a) Real-time anneal curve profile (98°C-75°C, ramping at 0.05°C/s), (b) A derivative plot of the anneal curve data, showing maximum rate of change over time, to give a defined anneal  $T_m$  value.

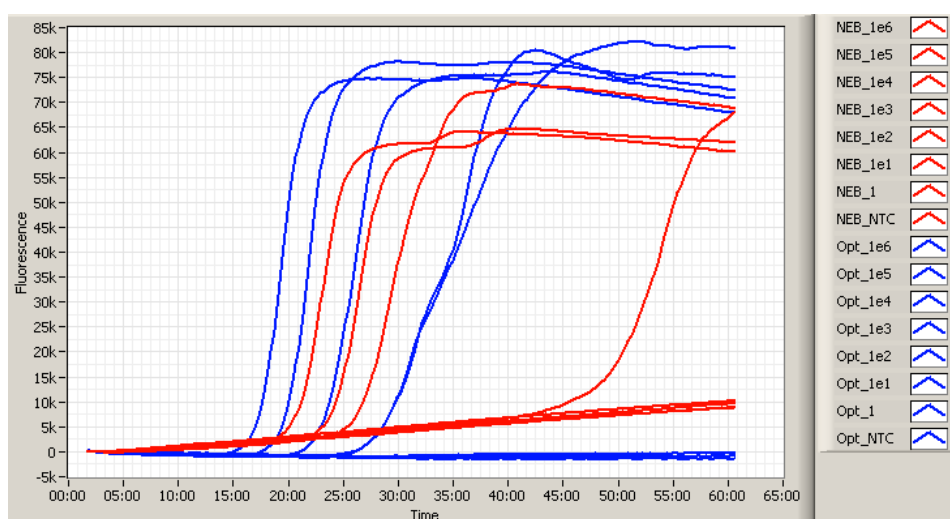
	Amplification time (min)	Anneal temperature (°C)
1X	31:01	83.54
0.5X	22:53	83.05
0.25X	19:55	82.82
0.125X	18:33	82.68
0.0625X	17:58	82.65
0.03125X	17:32	82.65
0.015625X	17:24	no data

**Table 5.3**

A comparison of the amplification times in the presence of varying concentrations of Eva Green<sup>®</sup> fluorescent dye as reported in Figure 5.15b. The anneal temperatures are also shown, as reported in Figure 5.16b.

### LAMP Reaction Buffer:

A reaction buffer can have a significant effect on an enzymes activity. *B.st* LF Pol I (NEB, UK) is provided with a 10X ThermoPol reaction buffer. This buffer has not been optimised for LAMP, and as such, variations in the buffer components may greatly improve enzyme speed, sensitivity and specificity. An alternative 'iBuffer' was reported to be an optimised reagent for LAMP (OptiGene Ltd, UK). The two buffers were therefore directly compared.



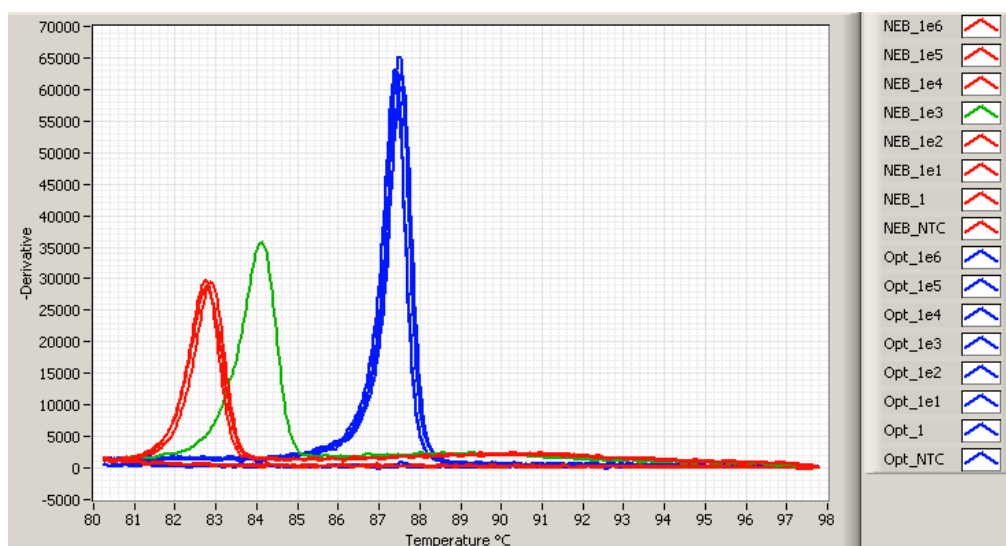
**Figure 5.18**

Isothermal amplification comparing the effect of reaction buffer on reaction sensitivity: ThermoPol Buffer (NEB, UK) (red), iBuffer (OptiGene Ltd. UK), (blue). LAMP reactions were run as reported in Chapter 2.6.5, using a dilution of *A.pe* gDNA template.

*B.st* LF Pol I amplified faster in the iBuffer and showed greater sensitivity than the ThermoPol buffer, enabling amplification from a further two template dilutions (down to  $1 \times 10^2$  copies) (Figure 5.18). The ability to run an anneal gradient post LAMP reaction shows the iBuffer also to show greater specificity (Figure 5.19 and Table 5.4). The ThermoPol buffer shows amplification of the  $1 \times 10^3$  dilution, but the anneal curve confirms this product to melt at a different temperature to the real amplicon ( $84.1^\circ\text{C}$  as opposed to the correct annealing temp of  $82.8^\circ\text{C}$ ), suggesting non-specific amplification has occurred.

A  $4.8^\circ\text{C}$  variation in anneal temperatures can be seen between genuine amplification products of iBuffer and ThermoPol buffer reactions. This variation is a result of the subtle buffer variations – increased  $(\text{NH}_4)_2\text{SO}_4$  and KCl concentrations, as well as an increased

Tris-HCl concentration. The baseline of the reactions containing ThermoPol buffer also shows a gradual increase in fluorescence over time, including the NTC reaction, further identifying this buffer unsuitable for real-time fluorescence detection in LAMP.



**Figure 5.19**

LAMP reaction anneal data, comparing the specificity of each reaction in either (i) ThermoPol Buffer (NEB, UK) (red), or (ii) iBuffer (OptiGene Ltd. UK), (blue). The green line highlights a false amplification product generated by the ThermoPol buffer at its template detection limit of  $1 \times 10^3$  copies of *A.pe* gDNA in this assay.

Buffer/copy no.	Amplification time (min)	Anneal temperature (°C)
ThermoPol_1e6	21:47	82.83
ThermoPol_1e5	25:05	82.85
ThermoPol_1e4	27:53	82.78
ThermoPol_1e3	52:41	84.10
ThermoPol_1e2	-	-
ThermoPol_1e1	-	-
NTC	-	-
iBuffer_1e6	17:56	87.38
iBuffer_1e5	20:25	87.45
iBuffer_1e4	24:20	87.51
iBuffer_1e3	29:44	87.51
iBuffer_1e2	30:22	87.47
iBuffer_1e1	35:15	87.37
NTC	-	-

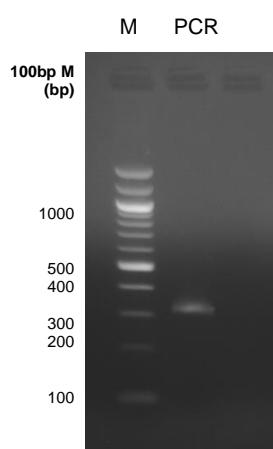
**Table 5.4**

Isothermal amplification times comparing effect of either ThermoPol Buffer (NEB, UK), or iBuffer (OptiGene Ltd. UK) on reaction sensitivity. LAMP reaction anneal data are also shown to compare the specificity of each reaction.

These results confirm that the iBuffer provides reaction conditions optimised for LAMP, enabling faster amplification speed, increased specificity and increased sensitivity. The iBuffer was therefore selected to be used for all further LAMP reactions.

#### *LAMP product confirmation:*

A LAMP reaction requires the specificity of 4 primers (F3, B3, FIP, BIP) for amplification. Notomi *et al.* (2000) have shown that amplification cannot occur unless all primers are present in the reaction. Two additional primers (LoopF and LoopB) can be included to provide extra priming sites, increasing the rate of amplification further. The specificity of 6 primers, targeting 8 specific regions, should provide enough confidence that the amplified product is derived from the template DNA and not due to false priming. To confirm this, 10ng *A.pe* gDNA was screened by PCR using the outer LAMP primers (F3 and B3) to generate a 310bp PCR fragment (Figure 5.20). The PCR amplicon was purified and sequenced using either the F3 or B3 primers. DNA sequence analysis confirmed the correct target had been amplified (Figure 5.20).



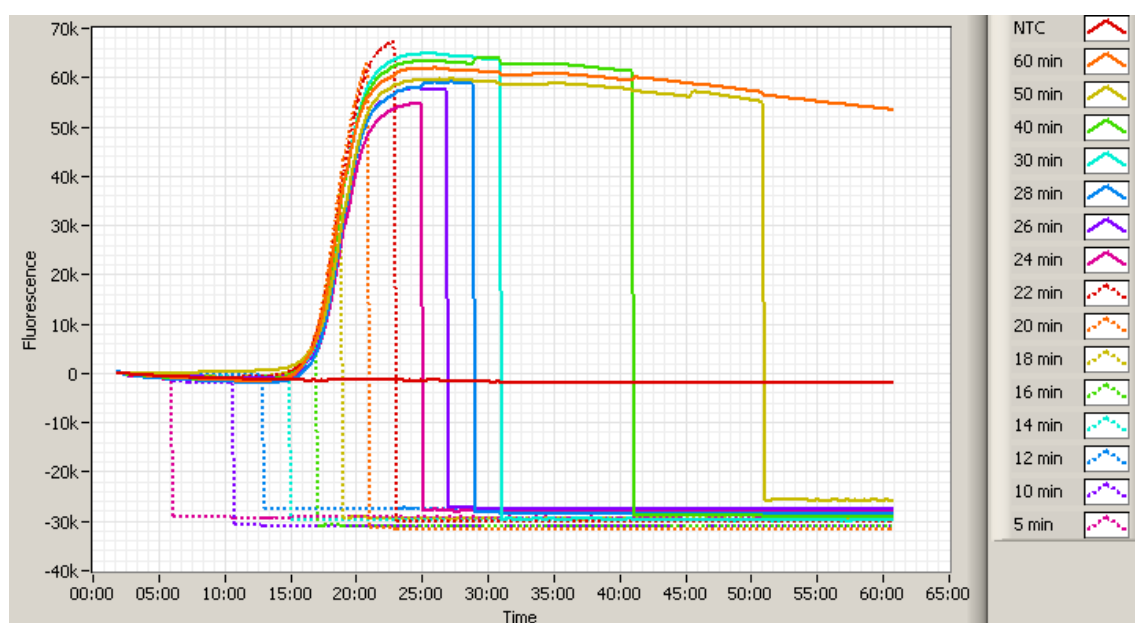
**Figure 5.20**

*A.pe* F3/B3 primer generated PCR amplicon, run out on a 2% (w/v) agarose gel. M: 100bp DNA ladder (NEB).

To confirm the amplicon, observed to melt at 87.4°C (+/-0.1°C), was the expected *A.pe* target DNA sequence defined by the LAMP primers, a restriction map of the amplicon was generated. The amplicon was defined by the FIP and BIP primers and showed a unique restriction endonuclease site for *Acc* I. When the LAMP product was digested with 10u *Acc* I the characteristic LAMP banding pattern was linearised, to a single 164bp, confirming specific amplification. This result had been previously confirmed by GeneSys Biotech Ltd.

### LAMP reaction characteristics:

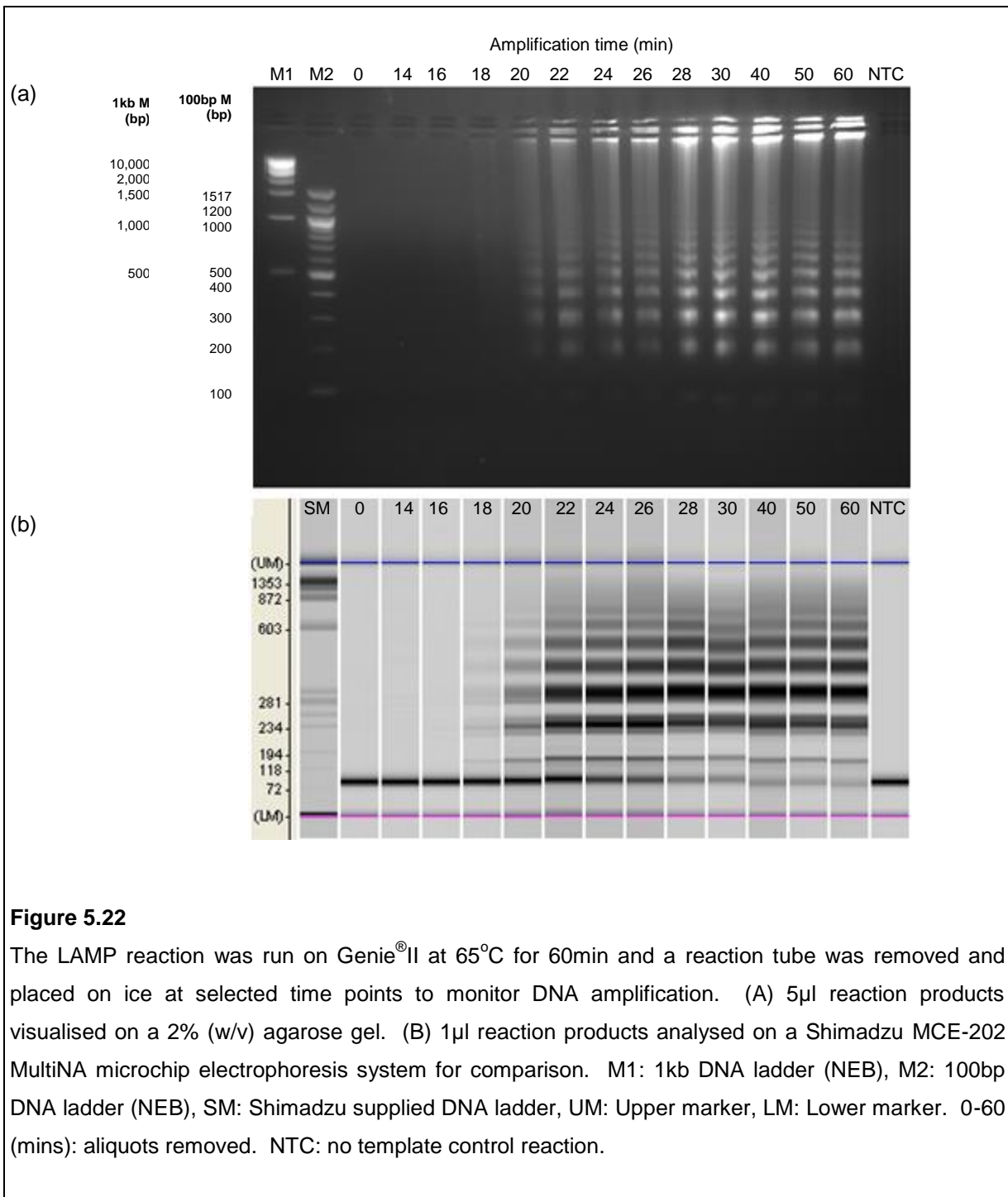
To gain a greater understanding of real-time LAMP characteristics, samples were taken at a variety of time points during the isothermal reaction to compare amplified products. Identical LAMP reactions containing  $1 \times 10^5$  copies *A.pe* gDNA were prepared, aliquoted into 16 tubes, and run on the Genie® II at 65°C. Real-time amplification was monitored, enabling samples to be removed as amplification progressed (Figure 5.21). Samples were placed at 85°C for 10min to stop the reaction and then frozen. After 60 min of amplification, all 16 reaction tubes had been removed, and 5µl (1/5<sup>th</sup> reaction) of each sample were visualised on a 2% (w/v) agarose gel to compare amplification products (Figure 5.22a). As the length of the reaction time increased, so too would the expected size of the stem-loop, inverted repeat products increase.

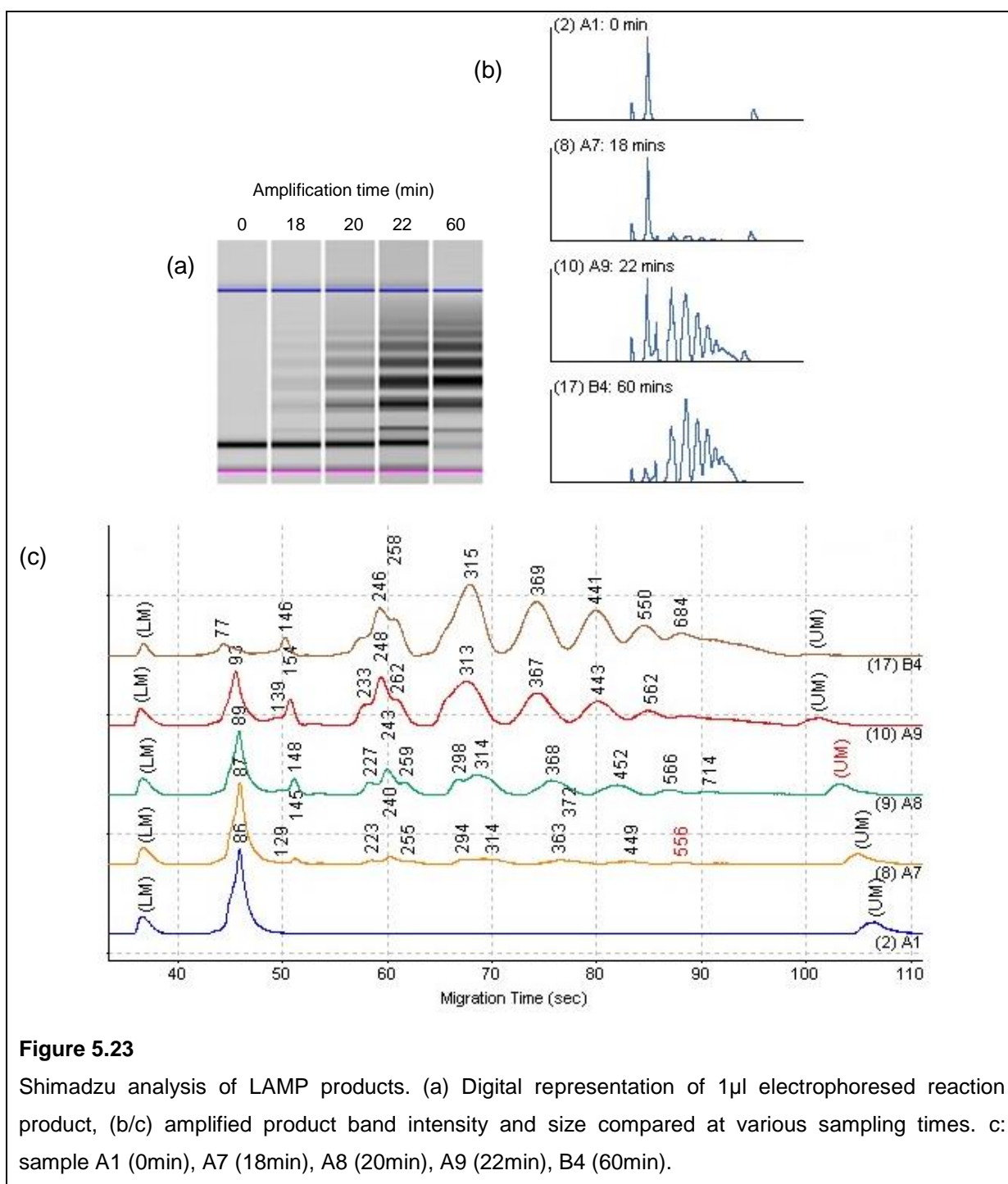


**Figure 5.21**

Isothermal amplification of  $1 \times 10^5$  copies *A.pe* gDNA target. Individual reactions were removed from the detection block at the times indicated, reported by a sudden drop in fluorescence.

The samples were further analysed using a Shimadzu MCE-202 MultiNA (Figure 5.23b). The microchip electrophoresis system provides detailed analysis of nucleic acids, including length of products (bp) and concentration (ng), when using standards stained with SYBR Gold1.





**Figure 5.23**

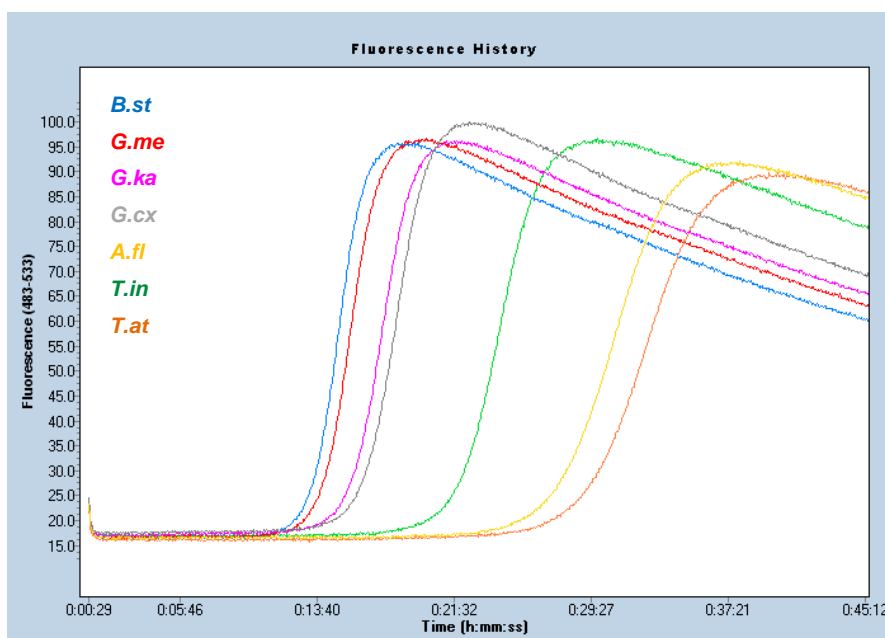
Shimadzu analysis of LAMP products. (a) Digital representation of 1µl electrophoresed reaction product, (b/c) amplified product band intensity and size compared at various sampling times. c: sample A1 (0min), A7 (18min), A8 (20min), A9 (22min), B4 (60min).

The analysed LAMP amplicons confirmed the products to increase in size with reaction time, as reported (Notomi *et al.*, 2000). With increased reaction time, the smaller amplicons can be seen to reduce as the *Bst* polymerase amplifies and strand displaces them into the larger, multimer repeating products (Figure 5.23a/b/c). The DNA analyser has a limited detection range up to the 1353bp upper marker, and therefore was only able to confirm product bands up to this size.

### Loop-mediated isothermal amplification

To compare the strand-displacement activity of the polymerases, a dilution of each enzyme was used in LAMP, using the reaction components previously optimised with *B.st* LF Pol I. The polymerases had already demonstrated the ability to extend a pre-annealed primer, and several, further displace a blocking primer, indicating strand-displacement activity was present. However, it was not known if the strand-displacement activity was sufficient for LAMP. The LAMP reaction requires dsDNA separation, by the polymerase, to first enable primer annealing. Multiple strand displacement sites are created throughout the reaction and it was not known if the increased strand displacement requirement of the polymerases would enable successful amplification.

The M13mp18 ssDNA primer extension assay demonstrated all polymerases showed optimal strand-displacement activity at 65°C (Figure 5.10 and 5.11), and therefore this was the defined LAMP reaction temperature for all comparisons. LAMP reactions were run following the method defined in Chapter 2.6.5, using a LightCycler®-480, enabling 96 reactions to be compared simultaneously. LAMP amplification results using the novel polymerases are reported in Figure 5.24 and 5.25.

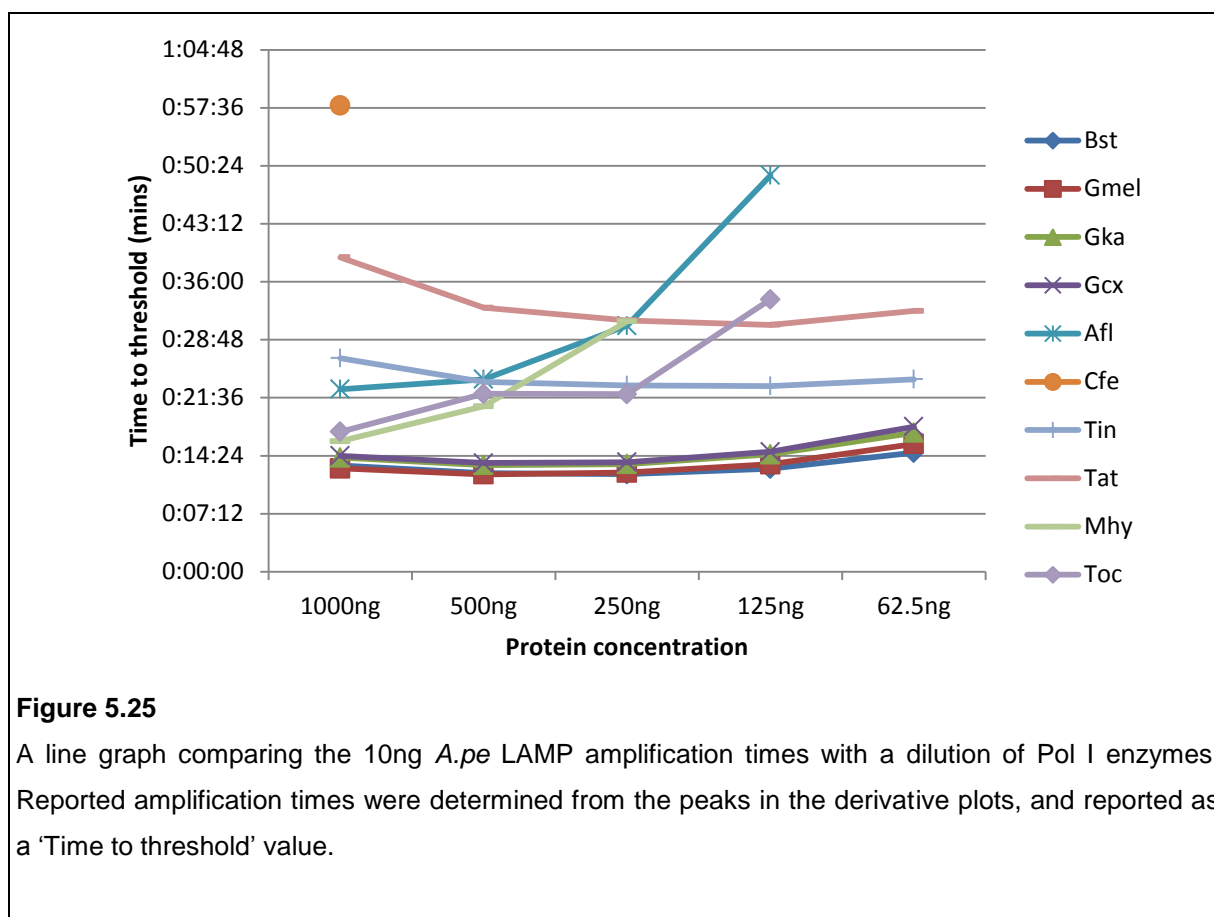


**Figure 5.24**

10ng *A.pe* LAMP reaction comparing polymerase activity: 8u *B.st* Pol I (blue), 62.5ng (approx. 8u *B.st* Pol I equivalent) *G.me* Pol I (red), 62.5ng *G.ka* Pol I (pink), 62.5ng *G.cx* Pol I (grey), 250ng (approx. 32u *B.st* Pol I equivalent) *A.fl* Pol I (yellow), 62.5ng *T.in* Pol I (green), 62.5ng *T.at* Pol I (orange).



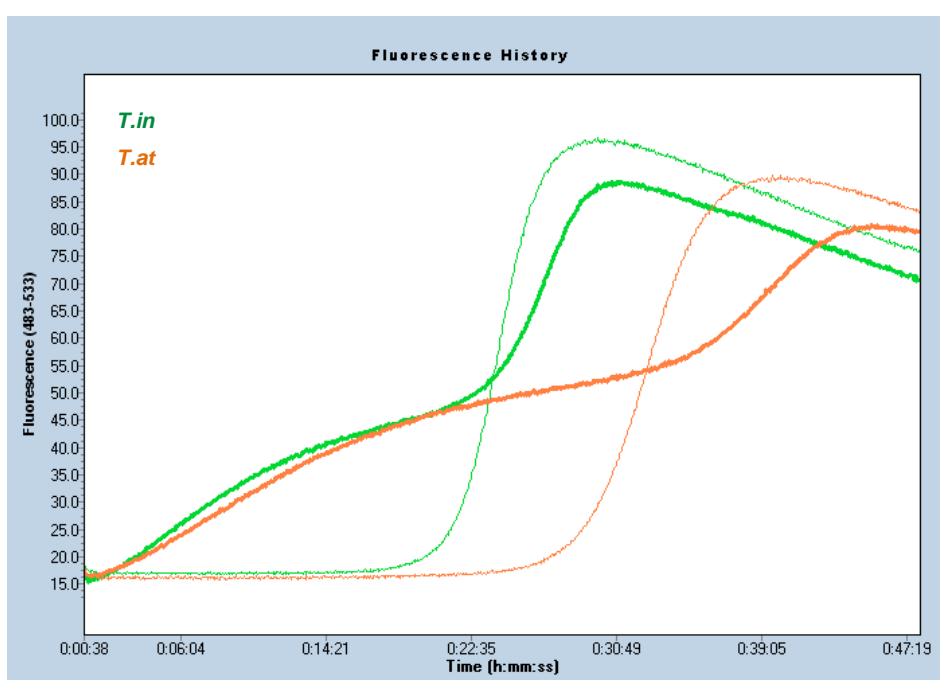
The polymerases demonstrating strand-displacement activity in the primer extension assay were further confirmed to be suitable for real-time LAMP amplification using the Eva Green<sup>®</sup> intercalating dye. Optimal concentrations of Pol I varied, with a general trend of increased enzyme resulting in faster amplification (Figure 5.25).



The *Geobacillus*-like polymerases showed little variation between amplification times, indicating each could be a direct alternative for *B.st* Pol I. Interestingly, several of the more thermostable polymerases were also shown to be suitable for LAMP. 62.5ng *T.in* Pol I (8u *B.st* Pol I equivalent) showed amplification 10min later than *B.st* Pol I, with little advantage observed by increasing the protein concentration. *T.at* Pol I was slower still, amplifying 20min later, using the equivalent protein concentration. The remaining thermostable polymerases required significantly higher concentrations of enzyme to achieve amplification, suggesting reduced enzyme activity, potentially as a result of a sub-optimal reaction buffer. *M.hy* Pol I required a 15-fold increase in protein concentration (1µg per reaction, the equivalent to 120u *B.st* Pol I) to achieve comparable amplification times to *B.st* Pol I. *C.fe* Pol I demonstrated the weakest activity in LAMP, amplifying at 58min with 1µg protein.

The M13mp18 ssDNA primer extension assay reported *K.o/* DNA polymerase to strand displace (Figure 5.11), but amplification did not occur in LAMP (within the 1hr reaction) with all protein concentrations tested. It was reasoned that this was because the strand-displacement activity was either not strong enough for LAMP, or the identified 3'-5' exonuclease activity was present, and substantially degrading the primers available for the LAMP reaction to occur.

Several amplification profiles showed an increased background of fluorescence that was more noticeable with increased protein concentrations (Figure 5.26, as observed by the bold amplification traces). The increasing fluorescent background appeared to be consistent with the use of a polymerase, predicted from its amino acid sequence, to possess 3'-5' exonuclease activity. This activity was suggested to be degrading the primers in the LAMP reaction and will be investigated further in Chapter 6.



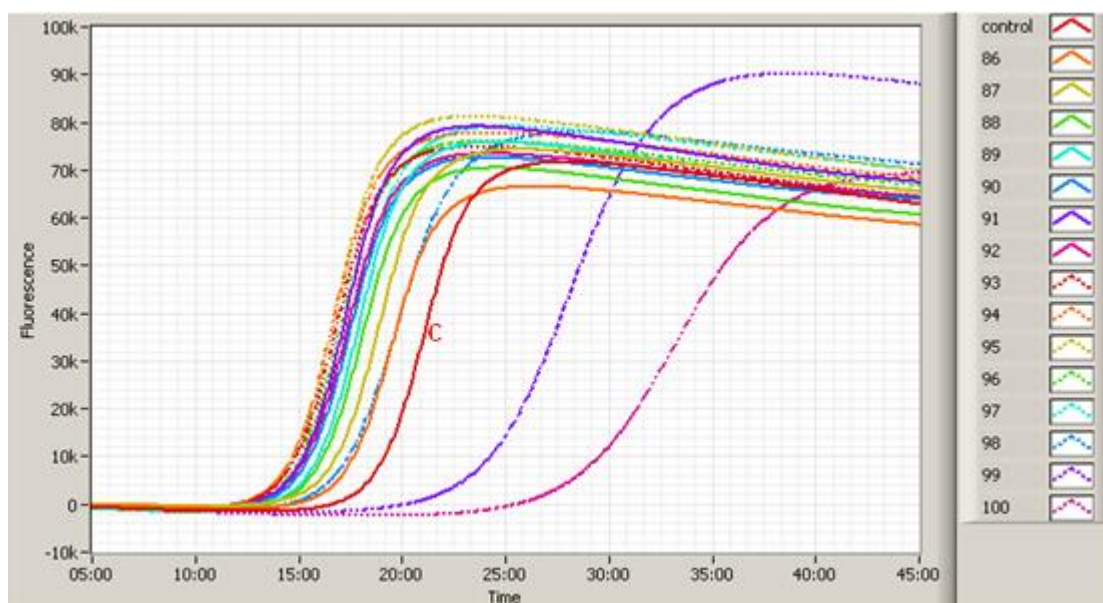
**Figure 5.26**

The increasing background fluorescence observed in a 10ng *A.pe* LAMP reaction, comparing 1000ng (bold) and 62.5ng (fine) Pol I activity. *T.in* Pol I (green), *T.at* Pol I (orange).

### 5.2.6 Heat-denaturing loop mediated isothermal amplification

The high thermostability of *T.in* LF DNA Pol I suggested the possibility of combining an initial heat step with the 65°C LAMP reaction. This 2-step, closed tube reaction, I have termed 'Heat-Denaturing LAMP' (HD-LAMP), may offer potential benefits over standard reaction conditions.

To test the effect of an initial template denaturation step on the speed of amplification in LAMP, identical reactions, uniquely including all reaction components (including *T.in* LF DNA Pol I), were incubated on a Genie®II for 5mins at a variety of temperatures. After incubation, the reaction temperature was reduced to 65°C to facilitate primer annealing for the LAMP reaction to occur. A reaction, without an initial heat-denaturing step, was included to serve as a control. Figure 5.27 shows the amplification profiles of the HD-LAMP reaction and clearly demonstrates the improved reaction speed when using an initial heat step. The non-HD, standard LAMP reaction can be seen to amplify at 19min 55sec (control 'C' - solid red line). HD-LAMP reactions, with initial heat steps between 86-98°C for 5mins, show faster amplification times; with the 95°C step demonstrating the fastest amplification time.



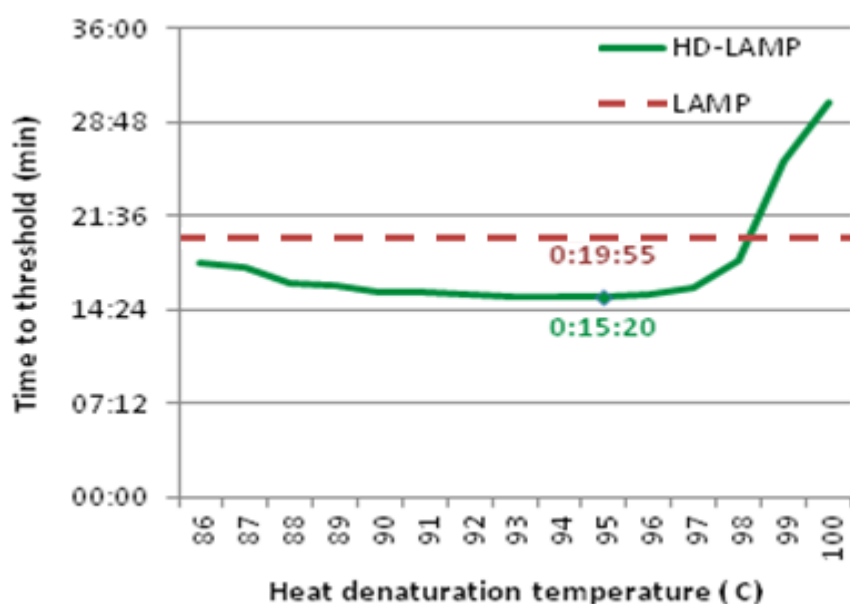
**Figure 5.27**

The effect of an initial heat-denaturation (HD) step on a closed tube LAMP reaction. LAMP reactions were run following the method in 2.6.5 with 125ng *T.in* Pol I. Control = no HD step and is further highlighted on the plot with a 'C'. The 5min HD steps, prior to LAMP were incubated at the reported temperatures.

This result further demonstrates the high thermostability of the *T.in* LF Pol I enzyme and the advantage that an initial heat step can convey to the LAMP reaction. Heat-denaturing steps above 95°C show an increasing amplification time with steps at 99 and 100°C showing later amplification times compared to the non-HD-LAMP control. This result suggests the enzyme is starting to denature above 95°C and confirms the protein melt analysis in Figure 5.6, where *T.in* LF Pol I is fully denatured above 100°C.

However, the 98°C HD-step (blue dotted line, Figure 5.27) shows a faster amplification time to that of the control reaction with without a heat-denaturing step (red solid line 'c', in Figure 5.27). Here, the benefit of template denaturation is therefore suggested to be greater than the reduction in enzyme activity through thermal denaturation.

An initial heat-denaturation step can be observed to reduce the amplification time of the LAMP reaction by 4mins 35sec when using a 95°C HD step, prior to LAMP (Figure 5.28).



**Figure 5.28**

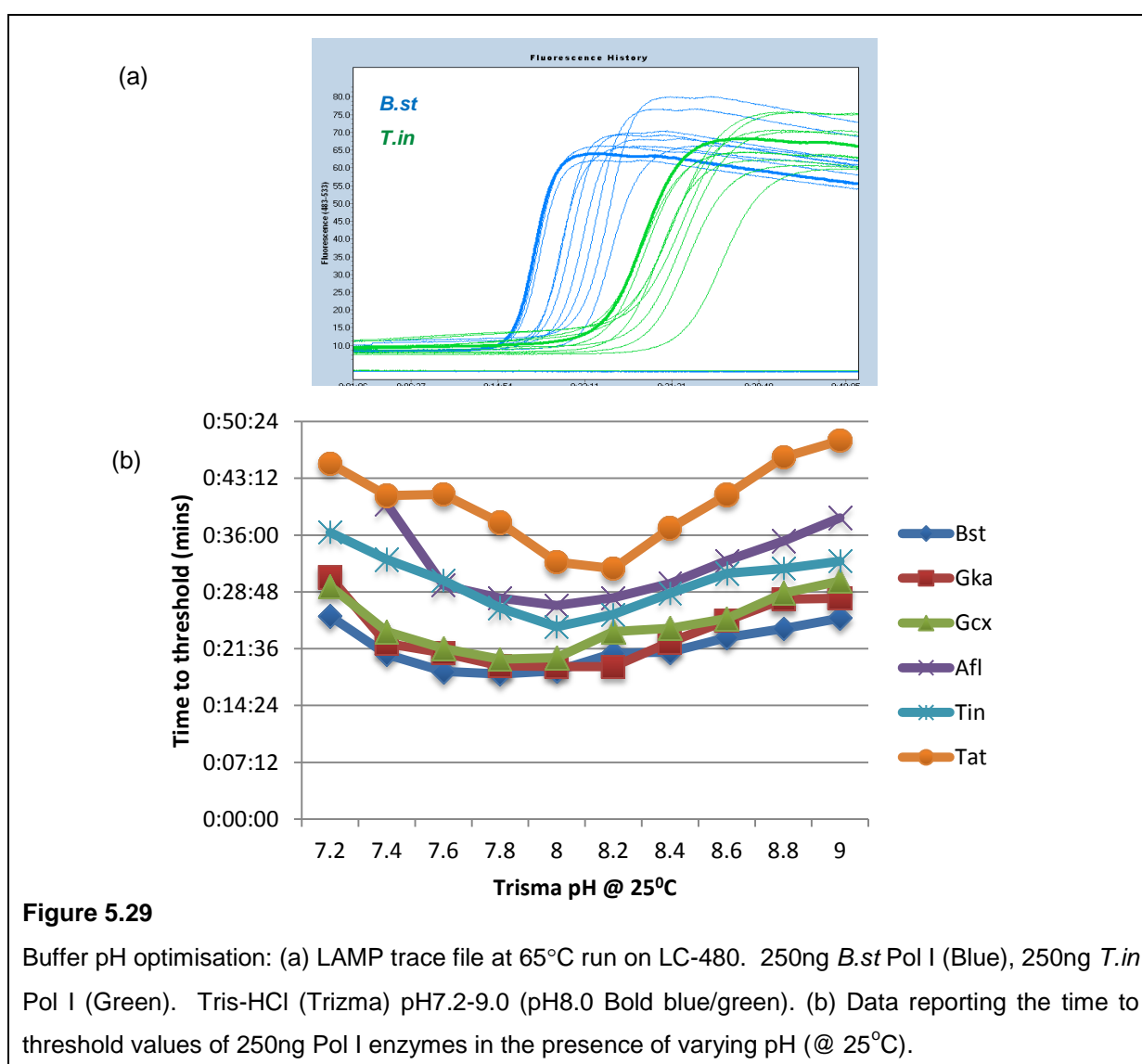
The effect of a heat-denaturation step prior to LAMP. The green line represents the amplification time of a LAMP reaction having had an initial 5min HD step at the displayed temperature (X-axis). The fastest reaction time is displayed and marked with a diamond. The red dotted line shows the reaction time of the control reaction with no HD step. Reported amplification times were determined from the peaks in the derivative plots, and reported as a 'Time to threshold' value.

## 5.2.7 Buffer optimisation

Having established several of the Pol I enzymes to be suitable for LAMP, the optimisation of individual buffer components were envisaged to improve the activity of the enzymes further. 250ng protein (equivalent to approximately 32u *B.st* Pol I), a concentration shown in Figure 5.25 to show sufficient LAMP activity from the Pol I enzymes, was used to compare the effects of components in a LAMP reaction. The following components were investigated using the method reported in 2.6.8.

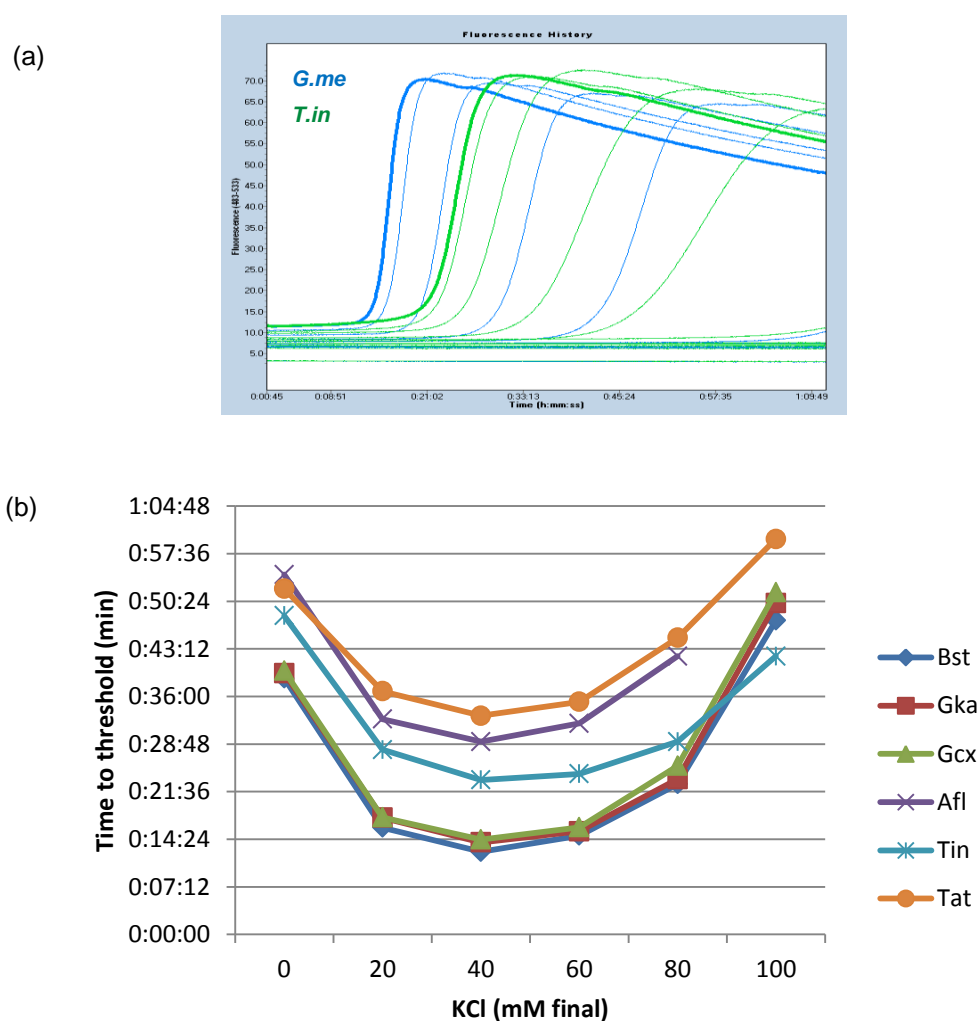
### Buffer pH:

All polymerases showed a preference for a reaction pH close to the iBuffer recipe (pH8.0), as indicated by a faster time to threshold value (Figure 5.29). Amplification times increased as the reaction pH became more alkaline or acidic than pH8.0.



### Salt tolerance:

The potassium chloride (KCl) concentration was also shown to have a significant effect on amplification times. All Pol I enzymes reported the fastest amplification activity with KCl concentrations between 20-60mM (Figure 5.30a/b). Away from these concentrations, the amplification times increased significantly, suggesting the environment was sub-optimal for each enzyme.

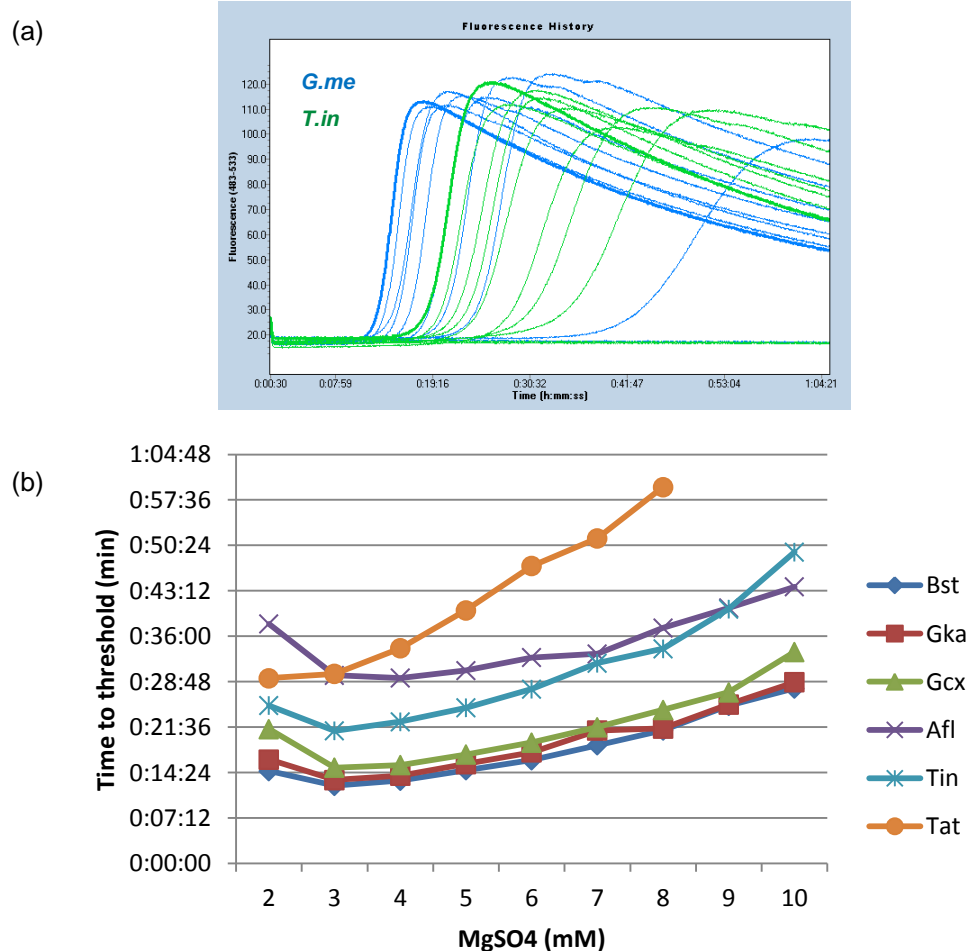


**Figure 5.30**

KCl buffer optimisation: (a) LAMP trace file at 65°C run on LC-480. 250ng *G.me* Pol I (Blue), 250ng *T.in* Pol I (Green). 0-200mM KCl final (20mM Bold Green, 40mM BOLD Blue). (b) Data reporting the time to threshold values of 20ng Pol I enzymes in the presence of varying KCl concentrations (final).

$MgSO_4$ :

Magnesium ions are required by the polymerase for nucleotidyl transfer reactions (Introduced in Chapter 1.1.2). Variations in  $MgSO_4$  concentrations were expected to have significant effects on polymerase activity. All Pol I enzymes showed the fastest amplification speed in the presence of 3-5mM  $MgSO_4$  final (Figure 5.31). The Pol I enzymes appeared to be tolerant to further increases in  $MgSO_4$ , but all showed reduced amplification times. 2mM  $MgSO_4$  increased amplification times for all polymerases, suggesting that the Magnesium ions were rate limiting to the reaction, with 1mM  $MgSO_4$  showing no amplification under 1h. This result further demonstrated the requirement for magnesium ions to facilitate DNA polymerase activity.

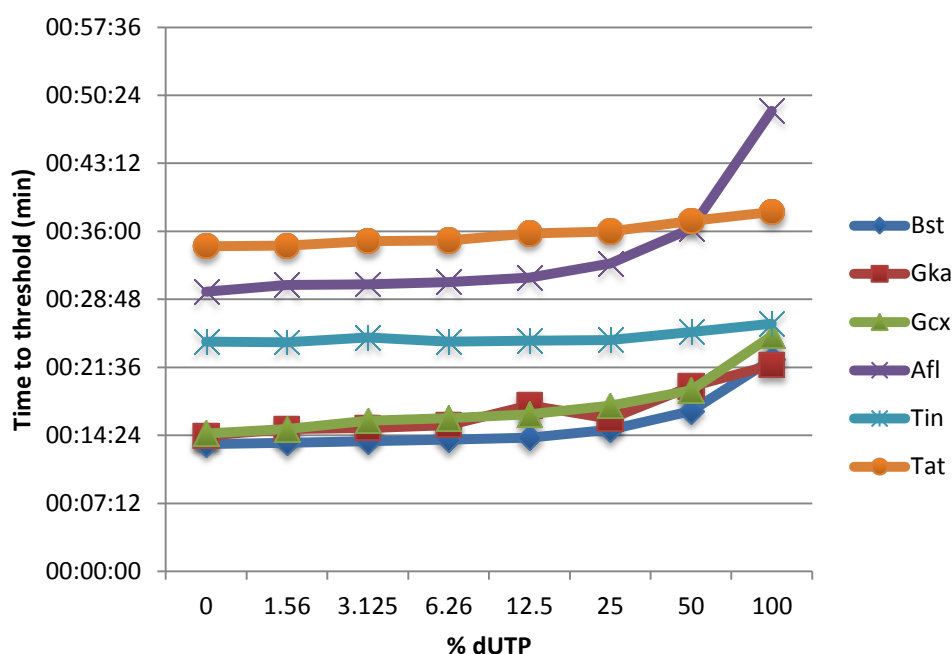


**Figure 5.31**

$MgSO_4$  buffer optimisation: : (a) LAMP trace file at 65°C run on LC-480. 250ng *G.me* Pol I (Blue) and 250ng *T.in* Pol I. 0-10mM  $MgSO_4$  final (4mM Bold Blue, 3mM Bold Green) (b) Data reporting the time to threshold values of 250ng Pol I enzymes in the presence of varying  $MgSO_4$  concentrations (final).

### 5.2.8 Incorporation of dUTP

The ability to incorporate modified bases may offer potential novel applications. LAMP reactions were run as described in Chapter 2.6.9. 250ng of each DNA Pol I were run in LAMP, with varying levels of dUTP in place of dTTP in the nucleotide reaction mix. The ability of each polymerase to incorporate dUTP can be seen in Figure 5.32.



**Figure 5.32**

Data reporting the time to threshold values of 250ng Pol I enzymes in the presence increasing dUTP concentrations (% final).

All polymerases were able to incorporate the alternative nucleotide into the growing DNA strands. The activity of the Pol I enzymes in the presence of a 100% dUTP containing mix was only slightly inhibitory to the majority of enzymes tested. *A.fl* Pol I showed a significant increase in amplification time, therefore a reduced ability of incorporation, in the presence of greater than a 25% dUTP containing mix.



## 5.2.9 Reverse transcription LAMP

To test whether the enzymes showed reverse transcriptase activity, 125ng each Pol I was compared in an RT-LAMP reaction. RT-LAMP primers were designed using LAMPDesigner (Premier Biosoft, USA) and run as described in Chapter 2.6.6, using 10ng MS2 RNA as the target sequence (Figure 5.33). The addition of 0.25U Avian Myeloblastosis Virus reverse transcriptase (AMV-RT) to a separate Pol I containing enzyme mix, was to be used as a control.

MS2 RNA was chosen as the RNA target because the bacteriophage does not include a DNA stage, therefore ensuring no background DNA would be present in the sample for the ssDNA LAMP primers to anneal. Amplification from the MS2 template can therefore only amplify from the RNA itself.

MS2\_FIP (F1c+F2) 5' -TCGCAAGCGAACCATCTACGTCCGTTGGTATAGACCTGAA-3'  
 MS2\_BIP (B1c+B2) 5' -CGATAGACTTATCGTCTGCATCCGAGATATGAATATAGCTCTGGTGG-3'  
 MS2\_F3 5' -TGTAAGGAGCCTGATATGAATATG-3'  
 MS2\_B3 5' -TAGTGTGAGCGGATACGAT-3'  
 MS2\_LoopF 5' -GCCAGACGCTGGTTGAT-3'  
 MS2\_LoopB 5' -GATCGCCTGGTGTGGAG-3'

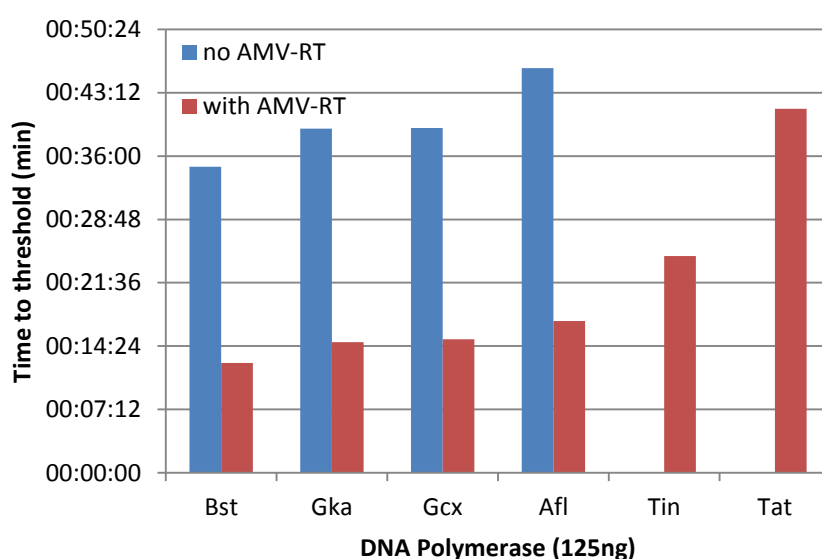


**Figure 5.33**

The MS2 cDNA target with the RT-LAMP primers and their locations.

The *Geobacillus*-like polymerases were shown to amplify the MS2 RNA target without a specific reverse transcriptase being required (Figure 5.34), but this activity was minimal when compared to the control reactions containing a specific RT enzyme. Amplification was observed between 34-49mins for these Pol I enzymes. In the absence of AMV-RT *T.in* and *T.at* Pol I did not show amplification in the presence of an RNA template within the 1hr reaction. Of the Pol I enzymes tested, only the *Geobacillus*-like pols demonstrated the ability to incorporate the bulky OH-containing uracil triphosphate of the RNA template, permitting cDNA synthesis to occur.

RT-LAMP reactions, with the addition of 0.25U AMV-RT, showed successful amplification from each Pol I tested. These results demonstrated the superior RNA-directed DNA polymerase activity of the AMV-RT enzyme. The AMV-RT synthesised complementary DNA (cDNA) significantly faster than the DNA Pol I enzymes. Once the cDNA was synthesised, it was available to the Pol I enzymes for DNA-directed DNA polymerase activity.



**Figure 5.34**

RT-LAMP time to threshold amplification: no AMV-RT containing LAMP reactions (blue), 0.25U AMV-RT containing LAMP reactions (red). Reaction were run at 65°C with 125ng each DNA Pol I.

RT-LAMP was also performed in the presence of an alternative metal ion. *Thermus thermophilus* (*T.th* Pol I) has previously been reported to possess efficient reverse transcriptase (RT) activity in the presence of  $MnCl_2$  yet limited activity with  $MgSO_4$  (Myers *et al.* 1991). To test the effect the RT-LAMP reactions were repeated as above, but contained the equivalent concentration of  $MnCl_2$ . No enzyme showed amplification from either the RT-

containing or RT-absent reactions in the presence of  $\text{MnCl}_2$ . This result indicated  $\text{MnCl}_2$  did not facilitate the DNA-directed DNA polymerase activity required for amplification.

It was concluded that the newly identified DNA Pol I enzymes in this report are not suitable for single enzyme RT-LAMP, and require the addition of a specific RT enzyme to facilitate a fast reaction.

#### **5.2.10 Polymerase unit concentrations**

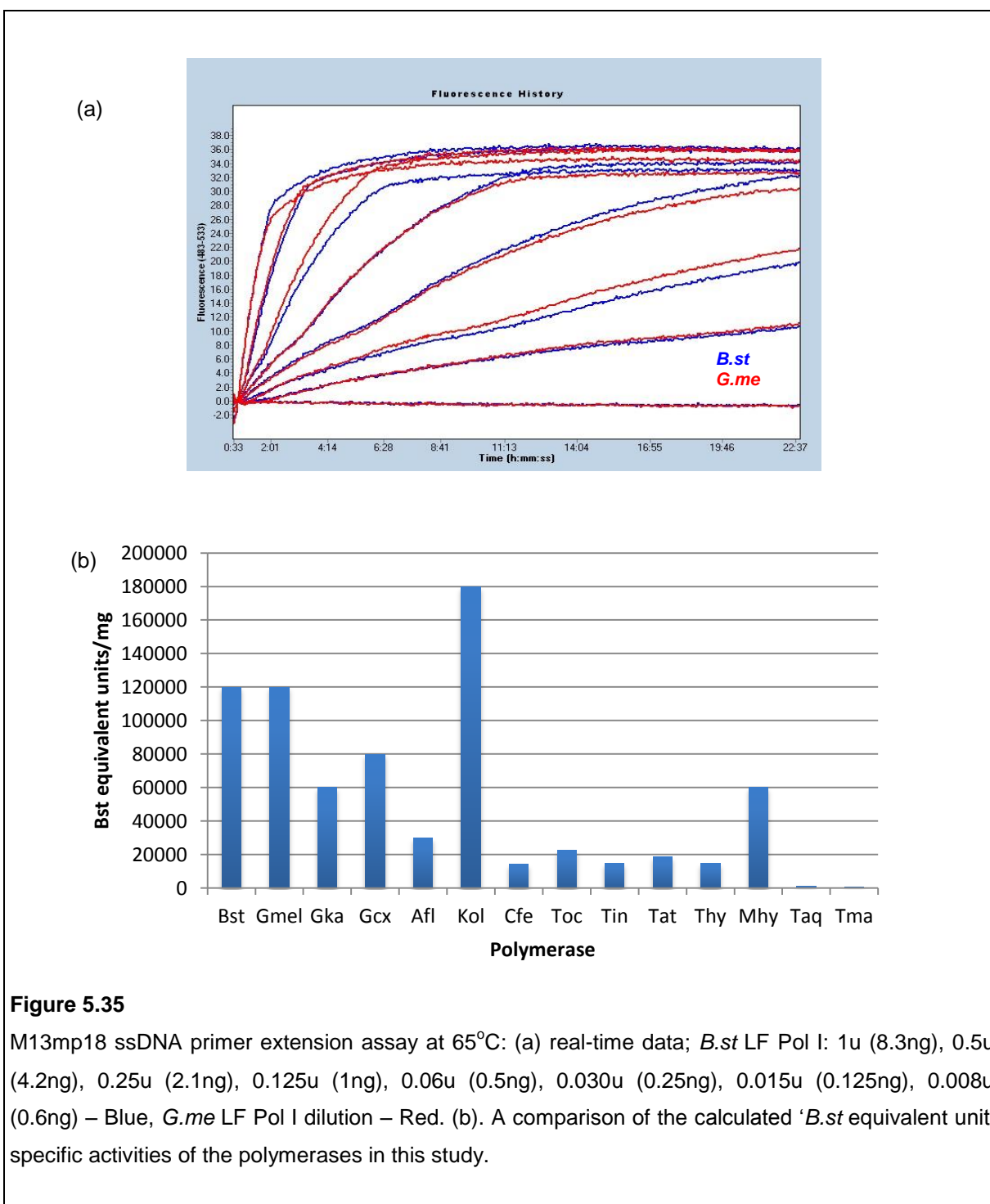
The M13 primer extension and LAMP assays had demonstrated the majority of enzymes to show strand-displacement activity at 65°C. These assays confirmed each polymerase to show 5'-3' DNA extension activity, but this activity had not yet been characterised.

The DNA extension activity of a DNA polymerase is classically defined using a radiolabelled dNTP incorporation assay, where one unit incorporates X nmol of deoxyribnucleoside-triphosphates into acid insoluble material in X minutes. The definitions are buffer and temperature specific to the enzyme in use, and therefore it is difficult to directly compare polymerases from a variety of organisms; with varying thermostabilities, and potentially unique buffer requirements. A safer, non-radiochemical approach was sought to enable direct activity comparisons of the Pol I enzymes in this report.

It was decided to define the reaction conditions to those suitable for a LAMP reaction, i.e. 65°C in the 1X iBuffer, to enable direct comparisons of the 5'-3' DNA polymerase activity. A fluorimetric primer extension assay was chosen to compare the activity of each Pol I. Activities could be directly compared to that of the gold-standard *B.st* LF Pol I under the same conditions, to assign a '*B.st* equivalent unit' value for each enzyme. Using this approach, polymerases with these defined units will therefore all incorporate the same number of nucleotides at the same rate.

Reactions were run as described in Chapter 2.6.10, in a 96-well plate, and monitored on a LightCycler®-480 at 65°C. A *B.st* Pol I dilution was used as a standard. *B.st* Pol I has a reported activity of 120,000 units/mg (NEB, UK); therefore 1U of enzyme contains approximately 8.3ng protein. A defined *B.st* equivalent unit value could then be assigned to the dilution of enzyme that matches the *B.st* Pol I extension rate. The specific activity for each polymerase under these conditions was then assigned as '*B.st* equivalent units/mg' protein for comparison.

Chapter 3 identified several enzymes with potential 3'-5' exonuclease 'proof-reading' activity. It was therefore necessary to use the phosphorothioated -47\_extension primer, described previously in Figure 5.8, to prevent degradation of the single-stranded oligonucleotide during the reaction, which may affect their calculated extension rates.



All polymerases were active as shown by an increase in fluorescence over time, confirming the 5'-3 DNA polymerase activity extending the M13mp18 ssDNA pre-annealed primer.

*G.me* LF Pol I showed an identical specific activity to that of *B.st* LF Pol I (Figure 5.35a). The remaining *Geobacillus*-like (*G.ka*, *G.cx* and *A.fl*) LF Pol I enzymes reported a reduced extension rate (Figure 5.35b). *K.ol* LF Pol I was identified as the enzyme with the highest 5'-3' DNA polymerase activity, with 180,000 *B.st* equivalent units/mg, under these experimental conditions. *C.fe*, *T.oc*, *T.in*, *T.at* and *M.hy* LF Pol I enzymes reported a lower level of activity compared to *B.st* LF Pol I, with *T.aq* and *T.ma* showing the lowest DNA polymerase activity under these conditions.

The high extension rate, in combination with strong strand-displacement activity, identify the *Geobacillus*-like polymerases to be of interest for isothermal reactions. The increased thermostability of the strand-displacing *T.in* and *T.at* LF Pol I enzymes may offer further potential advantages and will be explored further.

#### **5.2.11 Protein crystallisation**

BLAST<sub>p</sub> analysis of the *T.in* LF DNA Pol I sequence showed only 55% identity to the nearest enzyme in the database. The addition of crystal data for this unique enzyme would greatly improve the understanding of its structure and function, and may lead to the identification of key regions and motifs responsible for its ability to strand-displace.

Chapter 4.3.4 calculated the *T.in* Pol I enzyme stock to be 3.75mg/ml. This concentrated stock was used for crystallography studies. Purified *T.in* Pol I was screened using five commercial 96-well screening plates with crystallisation across two protein concentrations (3.75, and 1.875mg/ml). Screening plates were envisaged to cover a wide variety of conditions to provide a starting point for crystallisation optimisation. The plates were: (i) S1/S2 screen, (ii) Heavy/light screen, (iii) JCSG screen, (iv) PGA screen, and (v) Morpheous screen. Plates were incubated in a constant 20°C incubator and drops analysed each day for crystals. After 4 weeks, no crystals were observed. The polymerase storage buffer was noted to contain 50% glycerol. Although not expected to significantly affect the formation of crystals, the enzyme stock was dialysed against a no-glycerol containing storage buffer, to reduce its final concentration to approximately 10%. It was hoped that the new enzyme stocks would allow the appropriate precipitation for the formation of *T.in* Pol I crystals. Plate screens were repeated, and incubated at 20°C and at 4°C to compare the effects of varying temperature.

Unfortunately, no crystals were recovered. The crystallisation of *T.in* Pol I was abandoned and further attention turned to the engineering of the enzyme.

### 5.3 Discussion

Purified polymerases were characterised using a variety of assays, each in comparison to the current isothermal 'gold standard' enzyme, *B.st* LF Pol I. The polymerases were confirmed to be free from contaminating nucleases and *E.coli* genomic DNA, both of which are potentially carried through from the host expression strain.

A fluorescent thermal shift assay characterised the unique protein denaturation profile of each Pol I, assigning the enzymes into thermostable and hyperthermostable groups. The reaction environment can have a significant effect on the denaturation profile of an enzyme (Licata *et al.*, 2003). All reactions were therefore determined in the isothermal reaction buffer used for subsequent strand-displacement characterisation assays. The moderately thermostable *B.st*, *G.me*, *G.ka*, *G.cx*, *A.fl*, *K.ol*, and *T.oc* LF Pol I enzymes were shown to denature between 65-80°C. The thermostable *C.fe*, *M.hy*, *T.hy*, *T.in*, and *T.at* LF Pol I enzymes showed denaturation profiles between 80-100°C, with the hyperthermostable, *T.aq* and *T.ma* LF DNA Pol I enzymes denaturing at 104°C.

All polymerases showed a thermostability value several degrees higher than the optimal growth temperature of the respective host organism. This is not an unexpected result. Table 3.2, in Chapter 3, identified the wide range of growth temperatures reported for each organism used in this study. An organism's essential replicative proteins must be active at the upper limit of its growth range, to permit cell proliferation.

The denaturation profiles of each polymerase demonstrated the temperature at which the enzyme begins to unfold (initial increase in fluorescence), the maximal rate of unfolding (the peak value of fluorescence), and the temperature at which the enzyme is fully denatured (fluorescence has returned to baseline, i.e. dF/dT on the derivative plot). The majority of polymerases were shown to be fully denatured by the corresponding upper growth limit reported. Several enzymes, however, showed polymerase thermostability significantly higher than the upper limit for growth. *Thermodesulfatator* organisms have a reported upper growth limit of 75°C for *T.in* (Moussard *et al.*, 2010) and 80°C for *T.at* (Alain *et al.*, 2004), yet their LF DNA Pol I started to unfold at approximately 80°C, with full denaturation by approximately 100°C.

The reaction conditions for polymerase activity comparisons were constrained by the same parameters defining all DNA amplification methods, including PCR. DNA polymerases require an oligonucleotide primer from which to extend. If this primer is prevented from

annealing to the template DNA, due to a reaction temperature higher than the primer annealing temperature, extension will not occur. For this reason, the primers required for DNA extension assays were designed to anneal at least 5°C above the highest assay temperature. Having defined the upper temperature denaturation limit of each polymerase, characterisation assays could be performed up to this limit in the knowledge the enzyme will be available to the reaction.

The DNA Pol I enzymes were first tested for their ability to strand-displace a pre-annealed blocking oligonucleotide primer on a single-stranded circular template. The use of circular M13mp18 ssDNA as a template for the extension assay enabled two fundamental activities to be observed simultaneously: (i) 5'-3' DNA polymerase activity, and (ii) strand-displacement activity. When a polymerase extends from the primer and reaches the 5' end of a blocking primer there are three possible outcomes: (i) polymerisation stops, (ii) the primer upstream of polymerisation is degraded, if 5'-3' exonuclease activity is present, or (iii) the primer-template duplex in the path of polymerisation is unwound, and displaced as ssDNA, allowing polymerisation to continue (Hamilton *et al.*, 2001).

Of the thirteen LF DNA Pol I enzymes characterised in this study, all demonstrated 5'-3' DNA polymerase activity. Ten were further shown to generate high molecular weight DNA products, demonstrating strand-displacing, rolling-circle, isothermal amplification. This result confirmed the common reaction buffer was not inhibiting DNA polymerase activity, providing sufficient conditions, including available Mg ions, for nucleotidyl transfer. The optimal reaction temperature for strand-displacement activity was shown to be 65°C, for all enzymes demonstrating activity. *T.hy*, *T.aq* and *T.ma* LF Pol I failed to displace the blocking primer, thereby preventing the extension of dsDNA products greater than 5kb. The addition of a 5kb blocking primer further indicated that the halt in polymerisation was due to the primer and not due to elements of DNA structure. To confirm whether the 6x Histidine tag on the N-terminal of *T.aq* LF Pol I was affecting strand-displacement activity, a commercial enzyme preparation (GeneSys Biotech Ltd.), with no His-tag, was compared and confirmed the previous result.

The strand-displacement activity of DNA polymerases has often been shown to be dependent on reactions temperatures. *T.li* (Vent) DNA polymerase is reported to demonstrate strand-displacement activity at 72°C but not at 55°C, even though polymerisation was functional at these temperatures (Kong *et al.*, 1993). In comparison, Perler *et al.* (1996) demonstrated that *B.st* and *B.ca* LF DNA Pol I show strand-displacement activity at a wide variety of temperatures, only restricted by the enzymes upper

thermostability limit. It has been suggested that differences in temperature optima between polymerase activity and strand-displacement activity reflect temperature-dependent conformational issues or that the two activities are not directly related to a single active site for all polymerases (Hamilton *et al.*, 2001). This does not explain the results observed for *T.hy*, *T.aq* and *T.ma* LF Pol I. The assay was performed over the temperature range required for growth by these organisms and therefore demonstrates either (i) the LF Pol I does not strand-displace, (ii) the reaction conditions did not permit strand-displacement activity, or (iii) the LF Pol I requires accessory proteins *in vivo* to achieve efficient strand-displacement activity. The assay conditions were shown to be suitable for all polymerases to report 5'-3' primer extension. DNA replication relies on the ability to gap fill and repair the Okazaki fragments, and this is a fundamental requirement for dsDNA life. It is therefore reasoned that accessory proteins may be further required for *T.hy*, *T.aq* and *T.ma* Pol I to efficiently strand-displace *in vitro*.

The polymerases were further compared in an isothermal DNA amplification method. LAMP was chosen due to its reported ease of use, reliability, specificity and sensitivity. Its ever increasing popularity can be deduced from the increasing number of papers demonstrating the method compared to alternative isothermal methods.

Under isothermal conditions, LAMP has the potential to be a highly portable detection system. However, end-point agarose gel visualisation of the amplified product is not suitable for portability. Additionally, the requirement to open the reaction tube at the end of a reaction will lead to contamination by the amplicon, potentially interfering with future reactions using the same primer set. LAMP can be visualised by a variety of different methods: by eye (turbidity), or a colour change using hydroxynaphthol blue or through the addition of double-stranded intercalating dyes such as ethidium bromide or SYBR<sup>®</sup> Green I. Real-time detection of amplified DNA, reported by the increase in fluorescence of a double-stranded DNA intercalating dye, offers the fastest and most sensitive detection method. Much like real-time PCR, a variety of intercalating dyes may be suitable depending on the inhibition levels of the DNA polymerase in use. Real-time amplification can be detected using standard RT-PCR instrumentation but the size and power requirement of existing machines removes the ability for LAMP to be a truly portable application. Genie<sup>®</sup>II (OptiGene Ltd.) is a compact yet highly sensitive heated fluorescence detection instrument and offers battery power operation enabling use away from the laboratory. The supplied software (GenieExplorer – OptiGene Ltd.) has also been optimised for the LAMP method enabling detailed amplification times to be reported. This instrument, in combination with the LightCycler-480 (LC-480, Roche, USA), was chosen for the real-time LAMP work.



After isothermal amplification had completed, an anneal curve can be run, which offers similar advantages to melting amplification products post qPCR. An anneal curve, generated by heating the reaction to 99°C and then monitoring the fluorescence as the reactions are cooled over a gradient (0.05°C/s) enables the exact annealing temperature of the LAMP amplicon to be resolved. The LAMP reaction amplifies identical sequences of the target amplicon, specified as the sequence between the FIP and BIP primer sites. As such, the LAMP reaction contains multiple copies of this identical amplicon, in a variety of tandem repeat and stem-loop structures (reviewed in Chapter 1.2.3). The double-stranded DNA amplicon is melted at a high temperature and re-anneals at a precise temperature, which will be unique to any sequence as determined by its specific GC content. An anneal curve can therefore be used to verify that the correct DNA target has been amplified and that the increase in fluorescence is not due to false priming of the LAMP primers, or through the generation of primer dimers.

The LAMP method was explored and optimised using *B.st* LF Pol I before the novel polymerases were tested. An effective primer design was required for all primer-based methods. There are important properties that define a suitable primer sequence, such as the GC content, melting temperature, self-complementarity and the free energy for hybridization at the 3'-end, as well as parameters related to secondary-structure formation. Many PCR primer design tools have been developed over the years and are widely used. However, many isothermal DNA amplification technologies, including LAMP, have more complex reaction processes and require a fine-tuned primer design. A LAMP specific primer design tool (LAMP Designer™, Premier Biosoft, USA) was chosen to enable a LAMP primer set to be designed using *Aeropyrum pernix* (*A.pe*) gDNA.

The fluorescent intercalating dye, EvaGreen®, was not found to be significantly inhibitory to *B.st* LF Pol I below a concentration of 0.125X, enabling real-time detection of the amplification products. Amplification products confirmed the real-time reaction, using both agarose gel electrophoresis and computationally, via a Shimadzu analyser, to confirm the formation of increasingly larger amplicons with reaction time. As the reaction time increased, the short, initialising target amplicons amplified into the large, multimer repeating, stem-looped amplicons characteristic of the LAMP reaction. The isothermal reaction buffer 'iBuffer', supplied by OptiGene Ltd, was shown to increase the specificity and sensitivity of amplification when compared to the ThermoPol buffer, supplied with *B.st* LF Pol I (NEB, UK). The iBuffer was therefore used for all future isothermal reaction comparisons. A comparison of the two buffer recipes, reported in Chapter 2.1.1, showed the iBuffer to contain an additional 20mM NH<sub>2</sub>SO<sub>4</sub> (final), 20mM KCl (final), and a reduced pH of 8.1. Subsequent

buffer optimisation, detailed in this chapter, using a variety of polymerases, confirmed the suitability of the iBuffer for LAMP.

Fluorescent LAMP amplification was indicated by a sigmoidal increase in fluorescence over time, similar to the typical amplification curve generated during a qPCR reaction. The reaction contains sufficient components to generate >25µg DNA from a small concentration of template DNA (Notomi *et al.*, 2000), and was confirmed by running an aliquot out on an agarose gel for visualisation. The real-time LAMP reaction is rate limited by the concentration of EvaGreen<sup>®</sup>, indicated by the maximal fluorescence value obtained until a fluorescence plateau is achieved. At this point, no more dye is available to bind to the ever increasing amplification product, as confirmed by the time-course gel based analysis. The high sensitivity of the instruments optics enable amplification to be achieved in real-time well before confirmation could be achieved using an end-point visualisation method. One interesting observation during real-time detection was a decreasing fluorescence value, once the maximal fluorescence value had been achieved. A reduction in fluorescence was reasoned to be either a masking effect of the dye, due to the ever increasing stem-looped DNA structures being amplified, or due to the destabilisation by the DNA:DNA interactions within the stem loops.

Nine out of the ten LF Pol I enzymes, showing strand-displacement activity in the M13mp18 ssDNA primer extension assay, were further shown to amplify the *A.pe* gDNA target in LAMP. Interestingly, *K.o*/ LF Pol I did not show activity in the LAMP reaction, yet clearly demonstrated strand-displacement activity in the blocked primer extension assay. This result will be explored further in Chapter 6.

An interesting characteristic, observed in a small selection of strand-displacing polymerases, was an increased background fluorescence prior to amplification. *T.in*, *T.at* and *K.o*/ LF DNA Pol I showed this increase, which was more obvious with higher protein concentrations. It was noted in Chapter 3.2.10 that these enzymes contain the amino acid residues to confer an active 3'-5' exonuclease domain. These results will be explored further in Chapter 6.

The reaction conditions for strand-displacement activity were measured using the isothermal LAMP method, with conditions enabling the fastest amplification time defined as optimal. Reaction buffer optimisation for LAMP is similar to that of PCR, and relies on achieving an environment suitable for effective primer/template annealing with conditions enabling high enzyme activity.

The enzymes showed a defined range of activity with pH, salt, and  $\text{MgSO}_4$  preferences. All polymerases reported optimal amplification activity between pH 7.8 and 8.0, with 30-60mM KCl, and in the presence of 3-5mM  $\text{MgSO}_4$ . The iBuffer, common to all previous isothermal characterisation reactions, covers all of the optimised conditions and therefore no further optimisation was required. At first glance this is a surprising result. The wide variety of organisms harbouring these novel enzymes grow optimally in diverse conditions, with salt, pH and temperature variations. However, it can be seen their DNA polymerases function under the same conditions *in vitro*, indicating their intracellular environments are similar. Extremophiles have evolved a variety of mechanisms to ensure their internal components can function in these extreme conditions, thereby enabling survival in their respective environments. Halophiles, for example, expend energy to exclude salt from their cytoplasm to avoid protein aggregation. They have been shown to accumulate organic compounds to act as osmoprotectants, preventing desiccation through osmotic stress. Specific potassium and sodium pumps further maintain a specific environment suitable for life (Oren *et al.*, 2002). Thermophiles have further adapted their enzymes to enable activity, with the common mechanisms summarised in Chapter 5.1.2. The highly conserved nature of the Family A DNA polymerase, functionally and structurally, confirm this common environment approach, as may the nature of the LAMP reaction itself, requiring fine environmental limits for successful amplification.

The effect of individual buffer components have been previously characterised in PCR.  $\text{K}^+$  ions can bind to the phosphate groups on the DNA backbone and stabilise the annealing of the primers to the template DNA.  $\text{NH}_4^+$ , which exists as both ammonium ion and ammonia under high temperature conditions, can interact with the hydrogen bonds between the DNA bases to destabilize the hydrogen bonds of mismatched bases. The combined effect of the two cations has been suggested to aid the high ratio of specific to nonspecific primer:template binding over a wide range of temperatures in PCR (Roux *et al.*, 2009).

Magnesium concentration is a crucial component for DNA polymerase activity and has been shown to greatly effect amplification of DNA in PCR. The role of Mg ions has been demonstrated to provide two functions: forming complexes with dNTPs to promote nucleotidyl transfer, and by DNA/DNA interactions increasing primer annealing. Components in the reaction (template, chelating agents such as EDTA, dNTPs and proteins) affect the amount of free magnesium present in the reaction. DNA polymerases are inactive in the absence of magnesium, while too much magnesium leads to higher levels of non-specific amplification and decreases enzyme fidelity (Kunkel *et al.*, 1990).

The MgSO<sub>4</sub> concentration in the iBuffer supplied 5mM to the reaction, which is more than 3 fold above the total dNTP concentration (0.4mM each dNTP). LAMP reactions showed the optimum concentration for all enzymes tested to fall between 3-4mM MgSO<sub>4</sub> (final), but showed a tolerance at 10mM, with a trend of increasing reaction time with increased magnesium. Lower concentrations (2mM final) were shown to increase reaction time, with 0 and 1mM concentrations failing to give amplification; presumably, insufficient free Mg<sup>2+</sup> ions were available to the polymerase for primer extension. The iBuffer, chosen for this study, was therefore confirmed to be a highly suitable buffer covering the wide range of LF Pol I enzymes. The higher MgSO<sub>4</sub> concentration of the buffer was shown to be sub-optimal, but was envisaged to be preferable in applications where Mg<sup>2+</sup> ions could be further chelated. The iBuffer was therefore a sufficient buffer for LAMP, and was not altered for further assays.

Having confirmed all LF Pol I enzymes were active, and the reaction buffer was suitable for their characterisation, additional properties of the enzymes were investigated.

The ability of each LF Pol I to incorporate alternative nucleotides was compared. Using a DNA template, the ability to incorporate deoxyuridine triphosphate (dUTP), in place of deoxythymidine triphosphate (dTTP) was investigated. LAMP reactions were run, using varying dUTP:dTTP ratios, and the amplification times compared. All polymerases demonstrated the ability to incorporate dUTP, with only a marginal increase in amplification time observed in the presence of a 100% dUTP mix, indicating a slight preference for dTTP.

The use of dUTP in place of dTTP can be envisaged for amplicon cross contamination prevention (Longo *et al.*, 1990). As previously mentioned, nucleic acid amplification reactions generate large quantities of the same amplicon, greater than 25µg per reaction in LAMP (Notomi *et al.*, 2000). The contamination of new reactions with trace amounts of these amplified products, called carry-over contamination, yields false-positive results. Longo *et al.* (1990) reported that the contamination can be controlled by (i) incorporating dUTP into the reaction products (by substituting dUTP for dTTP), (ii) treating all subsequent set-up reactions with uracil DNA glycosylase (UNG), or (iii) thermal inactivation of UNG prior to amplification. UNG cleaves the uracil base from the phosphodiester backbone of uracil-containing DNA, but has no effect on natural (i.e., thymine-containing) DNA. UNG does not react with dUTP, and is inactivated prior to the amplification step. With all polymerases demonstrating the ability to incorporate dUTP, with minimal inhibition to activity, this is an effective way to potentially minimise cross contamination.

The ability of each polymerase to use an RNA template, i.e to possess RNA-dependent DNA polymerase activity, was investigated. Bacteriophage MS2 ssRNA was chosen to be used as the template for a reverse transcriptase LAMP (RT-LAMP). It has an advantage over alternative RNAs as there is no DNA intermediate involved during MS2 replication (Zinder *et al.*, 1975). This ensured amplification was not due to an MS2 DNA contamination in the template. RT-LAMP reactions were compared with and without a specific reverse transcriptase, AMV-RT, commonly used for RT-PCR in complementary DNA (cDNA) synthesis. The *Geobacillus*-like LF Pol I (*B.st*, *G.me*, *G.ka*, *G.cx* and *A.fl*) demonstrated RNA-dependent DNA polymerase activity, using the MS2 RNA template. The use of an RNA template significantly increased amplification times (approximately 2.5-fold increase) compared to reactions using a specific RT enzyme. This result confirmed that the polymerases could accommodate the large OH-group of the dUTP within the active site binding pocket, permitting nucleotidyl transfer, albeit not as efficiently as the standard dTTP nucleotide. *T.in* and *T.at* LF Pol I solely required the addition of a specific reverse transcriptase for amplification, indicating the binding pocket is not flexible enough to accommodate the RNA base.

The use of  $MnCl_2$ , was also investigated in the RT-LAMP assay. *Thermus thermophilus* (*T.th*) DNA Pol I has reported RNA-dependent DNA polymerase activity exclusively in the presence of manganese ions (Myers *et al.*, 1991). AMV-RT has also been demonstrated to efficiently utilise  $MnCl_2$  (Life Technologies, AMV-RT user manual). No amplification was observed in reactions containing  $MnCl_2$ , indicating the polymerases tested require  $MgSO_4$  for efficient LAMP amplification. AMV-RT has a reported optimal activity at 43°C and is active up to 55°C. The AMV-RT-containing LAMP reaction was incubated at 65°C, which is optimal for DNA-dependent DNA polymerase activity yet is above the maximal temperature for the RT enzyme. This demonstrates that enough cDNA is synthesised by the AMV-RT enzyme in the short time taken to reach the reaction temperature. With only a low amount of cDNA available for the LAMP template, this result further demonstrates the sensitivity of the LAMP reaction.

The focus of this project was to identify novel polymerases for strand-displacement activity; however, a major activity of all polymerases is the ability to extend DNA from a primer. The amplification speed of the LAMP reaction is defined by the ability to strand-displace, and also the rate at which the enzyme extends along the template. As previously mentioned, traditional unit definitions for a polymerase rely solely on the extension activity, not taking into account the effect of strand-displacement. The unit definitions also require the use of hazardous radiochemical assays, and those assays vary between manufacturers, with each

enzyme in its own specific reaction buffer. It was therefore decided to use a non-radiolabelled fluorescent primer extension assay to compare all the polymerases in this report, using the gold standard *B.st* LF Pol I as a standard. In this way, a '*B.st*-equivalent unit' value could be assigned to each polymerase, enabling direct comparisons in units/mg. All enzymes compared directly to *B.st* LF Pol I would therefore incorporate the same number of nucleotides at the same rate. This would then identify the effect of each enzyme's ability to further strand-displace.

*B.st* LF Pol I was assigned a unit definition of 120,000 units/mg enzyme, as reported from the literature, using traditional radiolabelled primer extension assays (NEB, UK). A dilution of each enzyme was run in a fluorescent primer extension assay and compared directly to the activity of *B.st* LF Pol I. The *Geobacillus*-like LF Pol I enzymes were all shown to have high extension rates, with *G.me* LF Pol I showing an identical rate to the *B.st* enzyme. Alignments in Chapter 3.2.10 identified a single amino acid residue variation between *G.me* and *B.st* LF Pol I, and this difference was not located within a key motif, and therefore was not expected to significantly alter activity. This result also indicated that the 6x Histidine tag at the N-terminus of *G.me* LF Pol I had no effect on polymerase activity. *G.ka* and *G.cx* LF Pol I were shown to have a reduced extension rate compared to *B.st* and also showed a reduced amplification speed in LAMP. The LAMP reaction generates long, stem-looped amplicon repeats requiring significant extension activity from the polymerase. The reduced LAMP activity is therefore reasoned to be due to a reduced extension rate rather than a reduced ability to strand-displace. This result is further supported by *A.fl* LF Pol I, which has a further reduced extension rate, and further reduced LAMP amplification speed.

*K.ol* LF Pol I demonstrated the highest extension rate of 180,000 units/mg, yet was not reported to strand-displace in LAMP. The 16S RNA analysis in Chapter 3 suggested *Kosmotoga* to share homology to the *Thermotoga* species, indicating the enzyme may share a more closely related polymerase. Pol I sequence alignments (Chapter 3.2.10) verified this high identity, with the results in this chapter verifying neither enzyme to show strand-displacement activity in LAMP. *M.hy* LF Pol I was shown to have a comparable extension rate to *G.ka* LF Pol I (80,000 units/mg). *M.hy* LF Pol I further demonstrated strand-displacement activity in LAMP. However, significantly more enzyme was required in the reaction (16-fold compared to *G.ka* LF Pol I) to achieve a similar amplification time in LAMP, suggesting its strand-displacement activity was considerably weaker than that of the *Geobacillus*-like polymerases.

The highly thermostable enzymes (*C.fe*, *T.oc*, *T.hy*, *T.in*, *T.at* LF Pol I enzymes) showed a

reduced extension rate when compared to the *Geobacillus*-like LF Pol I. *T.aq* and *T.ma* demonstrated the lowest rates with >120 fold lower specific activity. This trend suggests enzyme activity decreases with increased thermostability. Thermophilic proteins are thought to be less flexible than mesophilic proteins at mesophilic temperatures, and therefore they may not be able to achieve the same degree of flexibility as their mesophilic counterparts, with a consequent reduction in catalytic efficiency (Danson *et al.*, 1996). The most interesting result was therefore demonstrated by *T.in* and *T.at* LF Pol I. With extension rates of 15,000 and 19,750 units/mg respectively, both polymerases were shown to amplify relatively fast in LAMP, compared to the remaining thermostable polymerases. *T.in* and *T.at* showed an 18-fold reduction in extension rate compared to *B.st* LF Pol I, but only a 1.4 and 2.1 fold reduced amplification time in LAMP. This result suggests *T.in* and *T.at* LF Pol I contain significantly greater strand-displacement activity relative to the other polymerases in this study.

Singh *et al.* (2007) suggested the presence of Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup> (*E.coli* sequence numbering) in Pol I enzymes to be a prerequisite for strand-displacement activity. DNA polymerases containing this SYR motif were therefore assumed to possess strand-displacement activity and potentially suitable for isothermal amplification reactions. The results from this chapter indicate this to be correct for the majority of enzymes tested. The sequence alignments, in Chapter 3.2.10, identify *T.hy* and *T.aq* LF Pol I to show the relevant residues as described by Singh *et al.* (2007). However, no strand-displacement activity could be identified from these enzymes. It is reasoned these highly thermostable enzymes do possess strand-displacement activity, as they are required as DNA repair enzymes, but activity is weak, and may require accessory proteins. The activity of these polymerases were investigated further, reported later in this thesis.

Thermostable single-stranded DNA binding proteins (SSBs) have been suggested to stimulate strand-displacement activity in *T.aq* Pol I *in-vivo* (Viguear *et al.*, 2001). The majority of LF Pol I enzymes in this study have shown strong strand-displacement activity, suitable for LAMP, without the need for these additional enzymes. This suggests further residues or motifs may be required to enable single-enzyme strand-displacement activity to the polymerase, over and above the residues identified by Singh *et al.* (2007). Identifying these possible regions will be the focus in Chapter 6.

A significant result in this chapter demonstrated the use of *T.in* Pol I for a modified LAMP reaction, newly termed heat-denaturing LAMP (HD-LAMP). Although not essential to the reaction, several reports have detailed the advantages of a pre-LAMP template denaturation

step. Aryan *et al.* (2010) reported a 200-fold increase in sensitivity when using a 96°C – 3min heat denaturation step for the detection of *Mycobacterium tuberculosis*. Suzuki *et al.* (2010) further reported the use of a heat denaturation step to increase the sensitivity of cytomegalovirus detection in LAMP. In these experiments the LAMP reaction components, including enzyme, had to be added after the high temperature step to prevent denaturation of the mesophilic *B.st* Pol I enzyme. *T.in* Pol I was demonstrated to remain active in LAMP, after 5min at 98°C, with only a partial loss of activity observed, and may be of significant interest as a future diagnostic enzyme.

In conclusion, this chapter has identified several polymerases to be suitable as a potential replacement for the *B.st* LF Pol I enzyme in LAMP. The high thermostability of *T.in* LF Pol I, in combination with its strong strand-displacement activity, may enable unique applications not currently available by the less thermostable *B.st* enzyme. Activity variations between polymerases will be investigated further in Chapter 6.



## 6. ENGINEERING

### 6.1 Introduction

Polymerase characterisation in Chapter 5 identified several enzymes that may serve as direct alternatives for *B.st* Pol I, or offer activities enabling isothermal amplification reactions in applications not currently achievable by a moderately thermostable enzyme.

Protein engineering of polymerases has been used to enhance a variety of their activities and characteristics, thereby broadening their application to biotechnology and increasing their commercial value. Such improvements include increasing thermostability, resistance to potential reaction inhibitors, improved ability to catalyse the incorporation of modified nucleotides, and enhanced processivity and fidelity. These enhancements are introduced either directly, by target site-directed mutagenesis, or randomly using a non-specific, random mutagenesis approach. Site-directed mutagenesis is a well-established technique (Saiki *et al.*, 1988) and can enable the introduction of point mutations using overlapping PCR primers (Horton *et al.*, 1989).

*T.in*, *T.at* and *K.ol* Pol I were identified to contain the essential residues, proposed by Derbyshire *et al.* (1988; 1991) to confer 3'-5' exonuclease activity. This activity may degrade the oligonucleotide primers, thus reducing their annealing temperature and resulting in false-priming to the template and the consequent amplification of non-specific products. It is proposed that the removal of this activity, by mutation of the conserved residues, may lead to faster amplification times in LAMP, providing more primers for the reaction.

#### 6.1.1 The Singh SYR strand-displacement motif

Joyce *et al.* (1995) first identified the Fingers subdomain that binds the single-stranded template across from and beyond the site of synthesis, and inferred to it a role in template fixation and template specificity. Ser<sup>769</sup> and Phe<sup>771</sup> are located on the O-helix of DNA Pol I as defined by Ollis *et al.* (1980). The O-helix is made up of O, O1 and O2 helices. This 3-helix bundle has been shown to share structural homology with the DNA-binding motif Mrf-2 (Yuan *et al.*, 1998). Singh *et al.* (2007) have suggested the motif is a functional unit that recognises or induces altered DNA structure and requires an aromatic residue at its apex for strand separation.

Sequence alignments show the three key residues, reported by Singh *et al.* (2007), to be

highly conserved among known strand-displacing polymerases. This conservation was shown in Figure 3.13 of Chapter 3.2.10. Singh *et al.* (2007) have therefore suggested the presence of Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup> (*E.co* Pol I numbering) in DNA pol I enzymes may be a prerequisite for strand displacement activity.

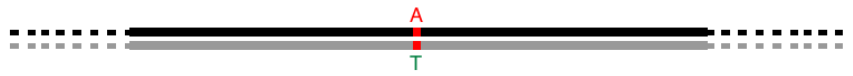
*T.hy* and *T.aq* Pol I were identified in Chapter 3 to possess the necessary residues for strand displacement, yet the results in Chapter 5 suggest the activity is not present, within the constraints of the reaction conditions. *T.ma* Pol I does not contain all the key residues, but as a Family A polymerase, extensively reported to require strand-displacement activity to repair Okazaki fragments, yet again, did not demonstrate this activity.

Singh *et al.* (2007) demonstrated that a mutation of these three key residues in *E.co* Pol I knocks out strand-displacement activity. They mutated each residue to Alanine but did not explore the effect of additional residues in these positions. It is assumed that mutations in the corresponding positions of *B.st* Pol I, and additional strand-displacing polymerases, confer the same reduction in activity. The effect of alternative residues in these positions is also not known and will be explored later in this chapter.

### **6.1.2 Molecular mutagenesis techniques**

Site-directed mutagenesis using overlap extension PCR reactions is a simple yet effective method to introduce mutations into a known gene (Horton *et al.*, 1989). The method is summarised in Figure 6.1 and will be used for all directed engineering in this chapter.

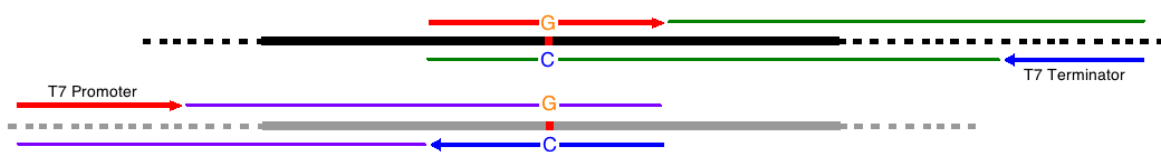
Stage 1: Identify DNA sequence to mutate



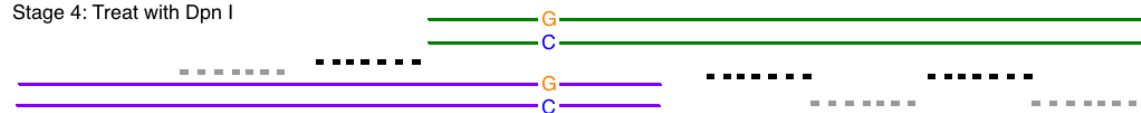
Stage 2: Design overlapping primer



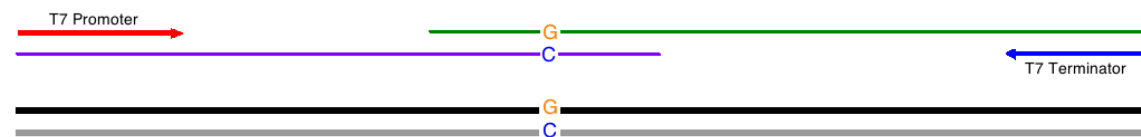
Stage 3: PCR amplify with specific primers



Stage 4: Treat with Dpn I



Stage 5: Second round PCR



**Figure 6.1**

Summary of the overlap extension PCR method. A site is selected from a known gene sequence for mutation (Stage 1). Primers are designed over that region, incorporating the mutated base, and consisting of at least 20bp either side, identical to the template sequence (Stage 2). Two PCR reactions are performed (i) using an upper primer (containing the ATG start codon) and the mutagenic lower primer (ii) a lower primer (containing the TAA stop codon) and the mutagenic upper primer (Stage 3). The reaction, preferably amplified from a plasmid DNA stock, is digested using *Dpn* I to remove the parental DNA template (Stage 4). The PCR products from Stage 3 are diluted and used as template for a PCR reaction using ATG and TAA encoding primers, to generate an amplicon containing the desired mutation.

Tang *et al.* (2012) reported a biased-codon primer method to replace the standard degenerate codon NNN that is often used to encode the 20 standard amino acids. An overly degenerate primer may lead to redundant codons and may remove the uneven distribution of amino acids in the constructed library (Tang *et al.*, 2012). This ‘small-intelligent’ primer consists of a mixture of four complementary primer pairs with a restricted degeneracy in each primer. The degeneracies are represented by NDT, VMA, ATG, and TGG at a stoichiometry of 12:6:1:1 respectively (Table 6.1). This improved codon degeneracy method will be used in this report to generate mutants where a wide selection of amino acid residues are required to be investigated at a single position.

Codon degeneracy	Amino acids represented
(12) NDT	N S I H R L Y C F D G V
(6) VMA	E A Q P K T
(1) ATG	M
(1) TGG	W

**Table 6.1**

Small-intelligent codon degeneracy, with represented amino acid residues.

### 6.1.3 Chapter overview

Further manipulation and engineering of the novel DNA polymerases identified in this thesis was expected to allow a greater understanding of the wide variations in activities identified. This chapter describes the attempts made to further investigate the roles of specific residues and motifs required for this strand-displacement activity.

## 6.2 Results

### 6.2.1 *G.me* LF DNA Polymerase I engineering

#### Cloning, expression and purification

DNA sequencing in Chapter 3 identified the DNA *polA* gene, encoded by *G.me*, to have 99% sequence identity to the commercially available *B.st* DNA *polA* (NEB, UK). Chapter 5 demonstrated the two enzymes to share common activities in a variety of assays and revealed the enzymes were superior to all other enzymes tested for use in LAMP. A DNA sequence alignment between *B.st* and *G.me* DNA *polA* identified 9bp sequence differences, but only one had an effect on the encoded amino acid; therefore there is only a single amino acid difference between the two Pol I enzymes (Figure 6.2). Ala<sup>433</sup> of *G.me* Pol I was mutated to the *Bst*-like Val<sup>433</sup> to provide a 100% identical amino acid sequence to *B.st* Pol I for future characterisation work.

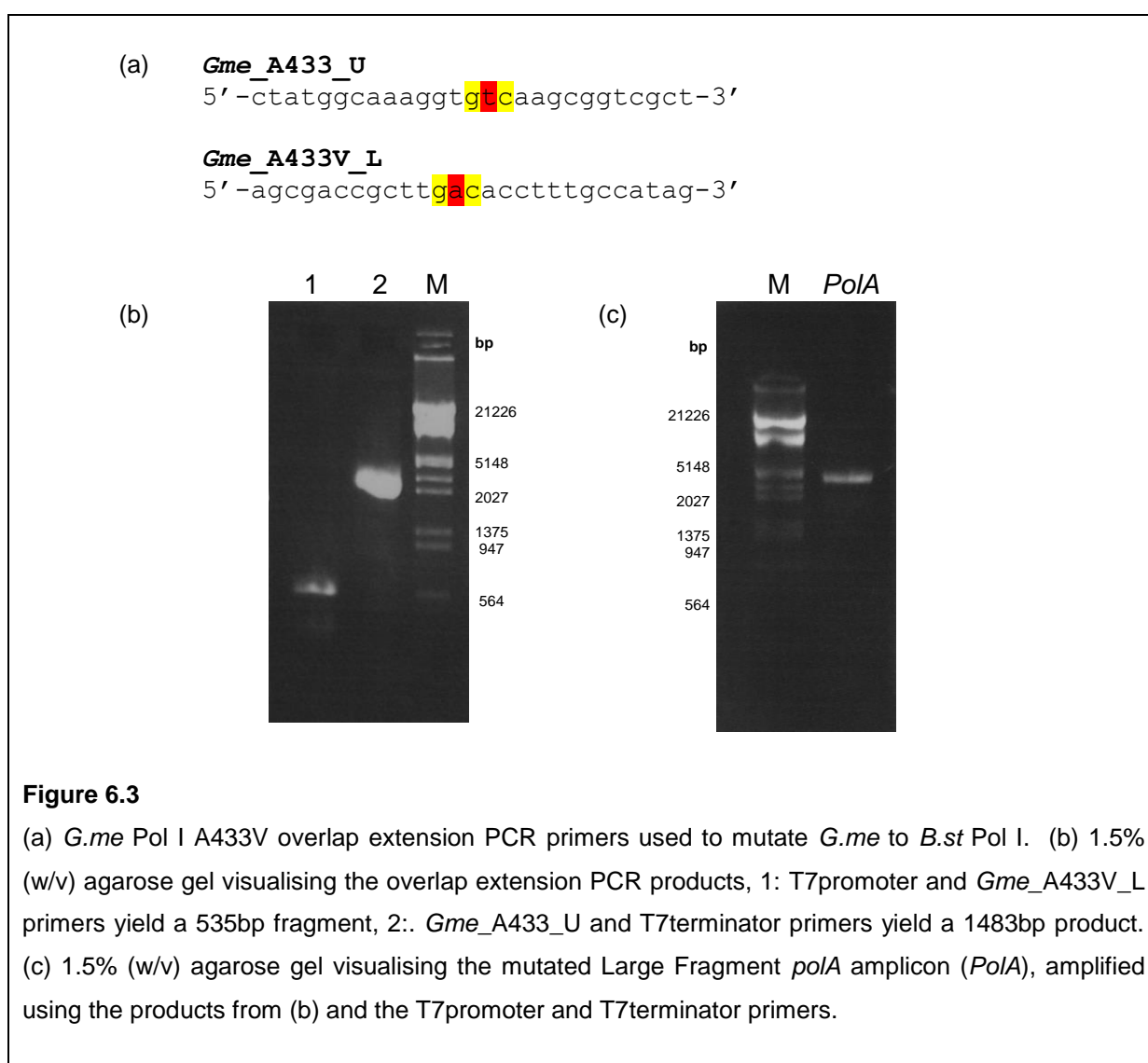
```
Gmel      TALADSQFLAWLADETKKKSMFDAKRAVVALKWKGIELRGVAFDLLLLAAYLLNPAQDAGD 120
Bst       TALADSQFLAWLADETKKKSMFDAKRAVVALKWKGIELRGVAFDLLLLAAYLLNPAQDAGD 120
          *****
          pos. 433
Gmel      IAAVAKMKQYEAVRSDEAVYGKGAKRSLPDEQTLAEHLVRKAAAIWALEQPFMDLNRNE 180
Bst       IAAVAKMKQYEAVRSDEAVYGKGVKRSLPDEQTLAEHLVRKAAAIWALEQPFMDLNRNE 180
          *****
Gmel      QDQLLTKLEQPLAAILAEMEFTGVNVDTKRLEQMGSELAEQLRAIEQRIYELAGQEFNIN 240
Bst       QDQLLTKLEQPLAAILAEMEFTGVNVDTKRLEQMGSELAEQLRAIEQRIYELAGQEFNIN 240
          *****
```

**Figure 6.2**

ClustalW sequence alignment of *B.st* and *G.me* Pol I. The varying alanine/valine residue is highlighted in yellow.

The construct containing the *G.me* Pol I clone was streaked onto an LB (Kan) agar plate, and plasmid from a single colony was purified to provide pET24a+HIS\_*G.me\_PolA* plasmid DNA. 10ng of the purified pDNA was used as template for subsequent PCR reactions. Primers were designed to mutate *G.me* Pol I A433V using the overlap extension PCR method (Figure 6.3a). Two high fidelity PCRs were run, the first containing primers T7promoter and *Gme\_A433V\_L* to give a 535bp product, and the second, *Gme\_A433\_U* and T7terminator to give a 1483bp product. Amplified DNA products were visualised on a 1.5% (w/v) agarose gel to confirm the correct fragment size (Figure 6.3b). PCR reactions were treated with *Dpn* I restriction endonuclease to digest the original template pDNA and

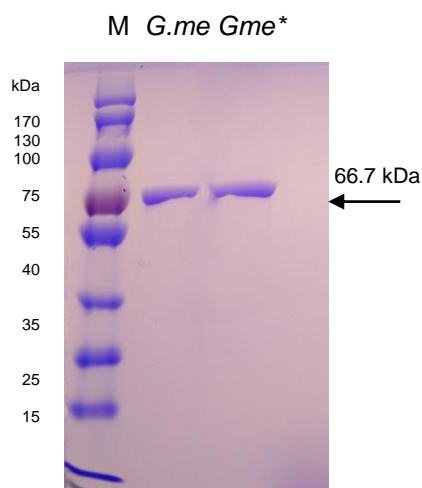
then heated to 85°C to denature the *Dpn* I enzyme. Each overlap PCR reaction was diluted 1/100 and then combined to be used as template for a second round of high fidelity PCR using the vector specific primers (T7promoter and T7terminator). The two PCR products provided overlapping sequences of 27bp (which incorporated the A433V mutation) to allow annealing for PCR. As such, all PCR amplicons derived from this template contained the 3'-5' exonuclease mutation. An aliquot of the PCR reaction was visualised on a 1.5% (w/v) agarose gel and then purified using a spin-column purification kit (Figure 6.3c).



The mutated PCR amplicon was *Nde* I and *Bam*H I digested, as described in Chapter 2.3.3, to allow directional cloning into the pET24a+HIS vector. The ligated clone was transformed into *E.coli* KRX(pRARE2) and colonies screened by PCR using the vector specific T7

primers. The amplicons were ExoSAP treated and DNA sequenced using the T7promoter primer to confirm the successful introduction of the A433V mutation. All recombinant clones showed the correct mutation and a single clone was over-expressed on a large scale and the recombinant enzyme purified as reported in Chapter 4.3. The mutant *G.me*\_A433V Pol I was re-labelled *G.me*\* Pol I.

Purified *G.me*\* Pol I was quantified using the Qubit<sup>®</sup> protein concentration method and 1µg was analysed by SDS-PAGE along with the wild-type (WT) *G.me* Pol I to confirm the purity of the sample (Figure 6.4).

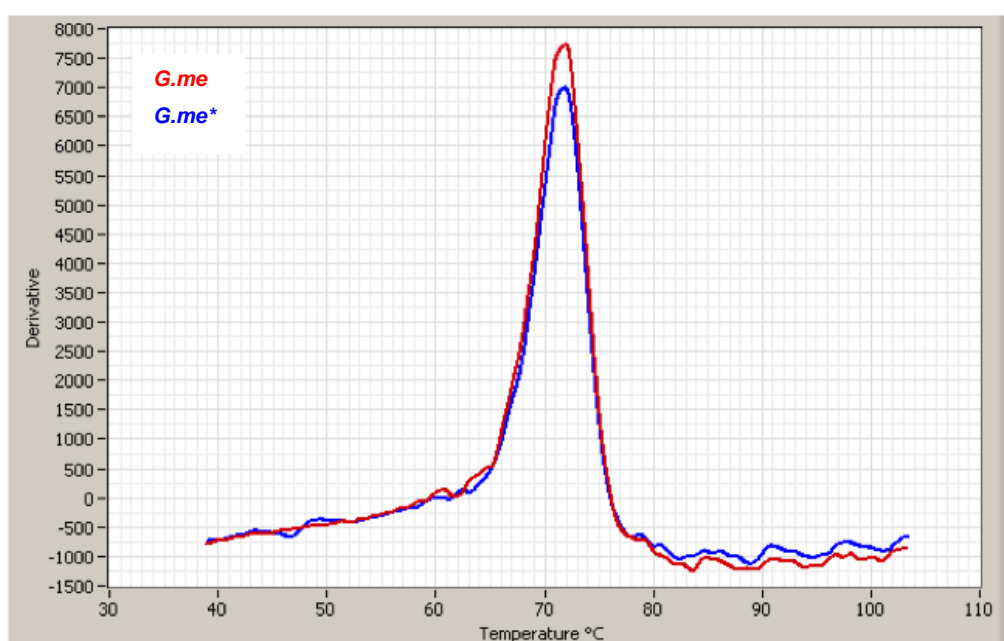


**Figure 6.4**

1µg purified protein electrophoresed on a 12% SDS-PAGE gel to compare the purity of *G.me* Pol I, and *G.me*\* Pol I. M: PageRuler protein standard. An arrow identifies the over expressed Pol I protein band.

## Characterisation

Purified *G.me*\* Pol I was characterised using the methods identified in Chapter 5. The SYPRO®-Orange thermal shift assay reported an identical protein denaturation profile to *G.me* Pol I, with a  $T_m$  of 71.8°C (Figure 6.5), indicating the mutation did not affect the proteins thermostability.



**Figure 6.5**

Thermal shift assay comparing the denaturation profile of 1µg *G.me* Pol I (Red) and *G.me*\* Pol I (Blue), in 1X iBuffer, in the presence of 10X SYPRO®-Orange, following the method in Chapter 2.6.3.

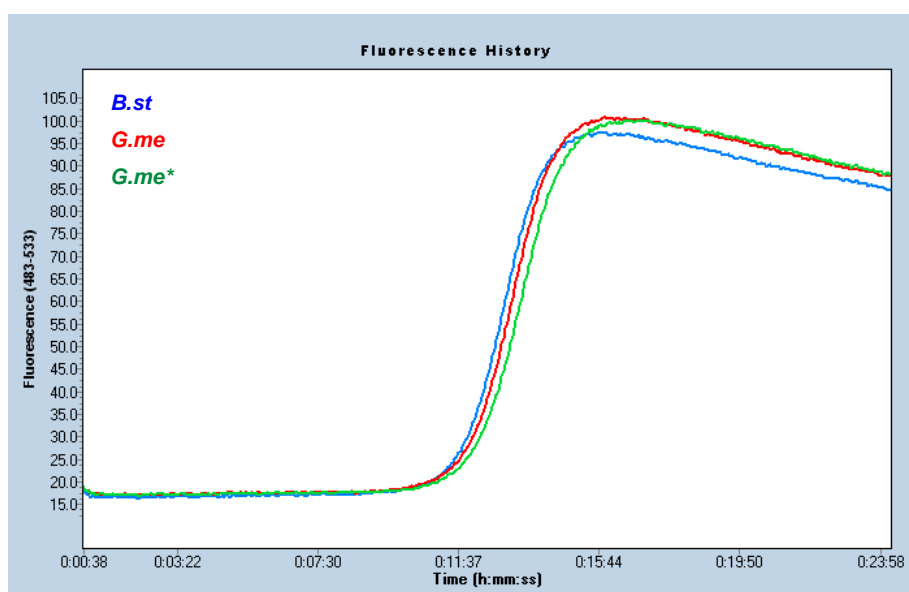
The M13mp18 ssDNA primer extension assay compared the activity of *G.me*\* Pol I to *B.st* and *G.me* Pol I. The results confirmed that all had identical extension rates of 120,000u/mg as expected (Table 6.2). LAMP assays further showed identical amplification times across a variety of template dilutions (Figure 6.6) and as such, *G.me*\* Pol I was demonstrated to be functionally equivalent to the commercially available, gold-standard, *B.st* Pol I. The *G.me*\* Pol I clone was therefore available for downstream mutation and chimera studies.



	Specific activity ( <i>B.st</i> units/mg)
<i>B.st</i>	120,000
<i>G.me</i>	120,000
<i>G.me</i> *	120,000

**Table 6.2**

Extension rates generated using the M13mp18 ssDNA primer extension assay, from Chapter 2.6.10, and calculated as *B.st* Pol I-equivalent units/mg.



**Figure 6.6**

LAMP amplification reaction: 10ng *A.pe* gDNA LAMP reaction incubated at 65°C and run on LC-480 with 125ng protein (16u *B.st* Pol I equivalent). *B.st* Pol I (Blue), *G.me* Pol I (Red), *G.me*\* Pol I (Green). The LAMP reaction was run following method 2.6.5.

## 6.2.2 3'-5' exonuclease knock out mutants

### Cloning, expression and purification

Sequencing of the DNA *polA* genes and their subsequent alignment identified several enzymes that contained a potentially active 3'-5' exonuclease domain (Figure 6.7). Asp<sup>355</sup> and Glu<sup>357</sup> of *E.coli* DNA Pol I have been previously reported to confer 3'-5' exonuclease activity (Derbyshire *et al.*, 1988). Their respective knock-out mutations, to alanine, have resulted in a significant loss of 3'-5' exonuclease activity. *T.in*, *T.at* and *K.ol* Pol I were chosen for mutational studies to characterise the effects of these mutations, with the aim to knock-out (KO) the 3'-5- exonuclease activity.

(a)

	3'-5' exonuclease motif		
	355	357	
<b>Tma</b>	EK LIEKLR ES PSFAIDLE	TSSSLDPFDCDIVGISVSFKPK EAYYIPLHHRN---	AQNLDEK 364
<b>Kol</b>	DEL FETLKGSDVISIDLE	TSSSLDPHEAKIVGISVAIKDGEGFYIPVNHESS--	DWQADEK 369
<b>Eco</b>	KAWIAKLEKAPVFADF	TETDSLDNISANLVGLSFAIEPGVAAYIPVAHDYLDAPDQISRE	399
<b>Tin</b>	KEIVEKAKGAEVVAIDLE	SDTKDPMRGKIVGVSLCFNPPKAYYF PFRHEGL-EAQKQLPW	373
<b>Tat</b>	HEILAKAREKGLVVIDLE	SNHIDPMRGKIVGVALCFEPPKAYYF PFRHEGL-EARKQLPW	373
Thy	KIDLEKLKKA EILSIKI	QGN-----LFSLTNPEELTVASDEK-EAYK-LSV	339
Bst	TEEMLADKAA--LVVEVMEENYH--	DAPIVGIALVNEHG-RFFMRPETALA-----	DS 355
Gme	TEEMLADKAA--LVVEVMEENYH--	DAPIVGIALVNEHG-RFFMRPETALA-----	DS 355
Bcx	TEEMLADKAA--LVVEVMEENYH--	DAPIVGIAVVNEHG-RFFLRPETALA-----	DP 66
Gka	TEEMLADKAA--LVVEVMEENYH--	DAPIVGIAVVNEHG-RFFLRPETALA-----	DP 357
Afl	DEHMLTKEGA--LVVEVL DENYH--	QAPIVG FALVNERG-HFFIPTDIALA-----	SS 361
Bfn	TDDL FASPAS--LVVEQLGDNYH--	EAPILGFSIVNEHG-AFFIPEETAVQ-----	SD 358
Toc	GRMMERVRAAGVLAVELKTDGRN	PMDAHLIGIGFSPSRGEGFYVPAEVLEK-----	SP 359
Cfe	-----	-----	
Taq	EEAPWPPPEG--AFVGFVLSRKEP	WADLLALAAARGGRVHRAPEPYKALR-----	343
Mth	EEAPWPPPE--AFLGYVLD RPQPM	WAE LKGLAGAWEGRVARGPARAKELA-----	345

(b)

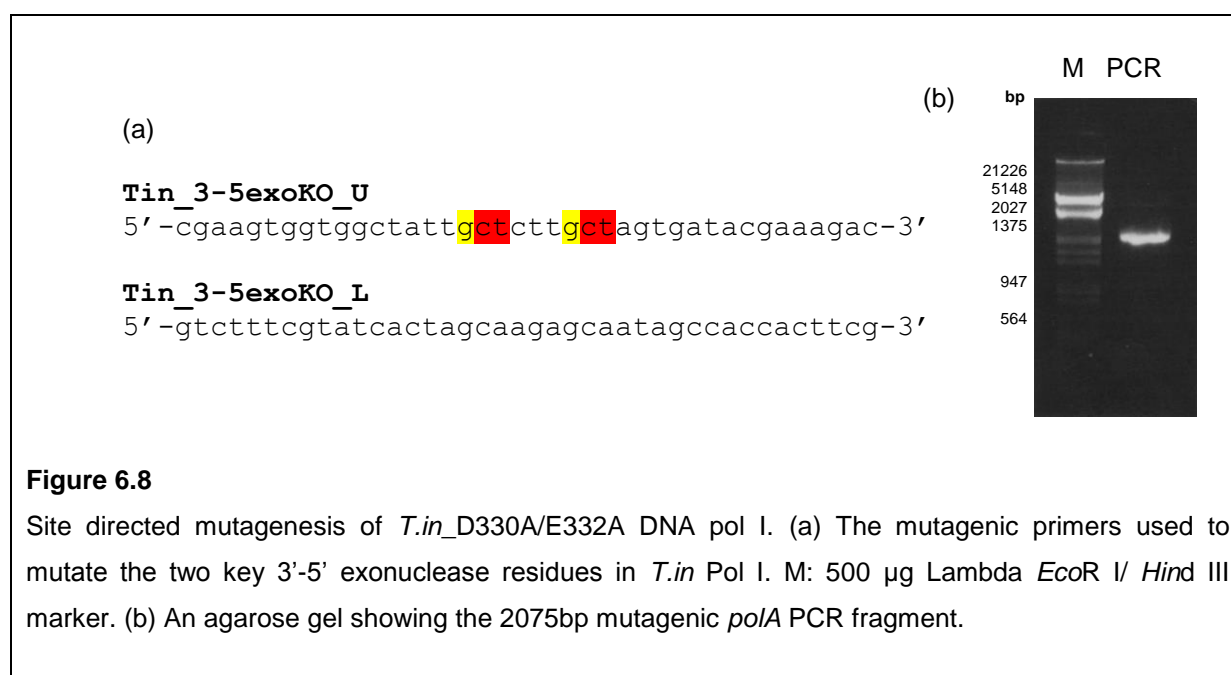


**Figure 6.7**

(a) ClustalW amino acid alignment of Pol I sequences. The key Asp and Glu residues, with numbering based on *E.co* Pol I, are highlighted in blue. (b) The position of the key residues within the Pol I sequence.

### *Thermodesulfatator indicus* 3'-5' exo KO

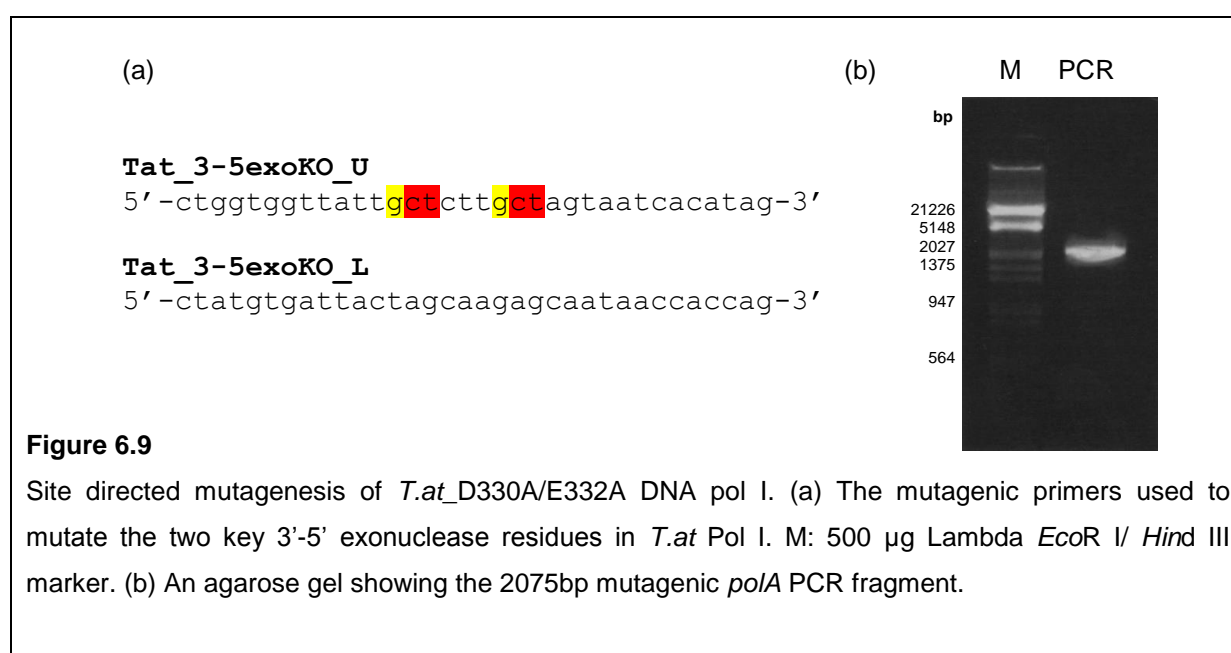
Plasmid DNA, containing the *T.in* DNA *polA* gene was purified for use as the DNA template for a site-directed mutagenesis PCR. Two overlapping PCR reactions, using a single primer pair, were run to generate the D330A and E332A mutations, following the method in 2.3.6. The first reaction contained primers T7promoter and Tin\_3-5exoKO\_L to give a 239bp product and the second, Tin\_3-5exoKO\_U and T7terminator, to give a 1483bp product. The mutagenic primer sequences are detailed in Figure 6.8a. Amplified DNA products were visualised on a 1.5% (w/v) agarose gel to confirm the correct fragment size. PCR reactions were treated with *Dpn* I restriction enzyme to remove the parental template vector DNA. Each overlap PCR reaction was diluted 1/100 and then combined to be used as template for a second round of high fidelity PCR using the vector specific primers (T7promoter and T7terminator). The two PCR products provided overlapping sequences of 40bp (which incorporated the D330A/E332A mutations) to allow annealing for PCR. An aliquot of the PCR reaction was visualised on a 1.5% (w/v) agarose gel (Figure 6.8b) and then purified using a spin column purification kit.



The mutated PCR amplicon was *Nde* I and *Bam*H I digested to allow directional cloning into the pET24a+HIS vector. The ligated clone was transformed into *E.coli* KRX(pRARE2). Five recombinant colonies were each screened by PCR using the vector specific T7 primers, and the amplicons were ExoSAP treated and DNA sequenced using the T7promoter primer to confirm the successful introduction of the D330A and E330A mutations.

### *Thermodesulfatator atlanticus* 3'-5' exo KO

Plasmid containing the *T.at polA* gene was purified for use as the DNA template for a site-directed mutagenesis PCR, as reported previously, to generate the *T.in polA* mutant. Again, two overlapping, high fidelity PCR reactions were run to generate the E430A and G432A mutations. The first reaction contained primers T7promoter and Tat\_3-5exoKO\_L to give a 237bp product, and the second, Tat\_3-5exoKO\_U and T7terminator to give a 1483bp product. The mutagenic primer sequences are detailed in Figure 6.9a. Amplified DNA products were visualised on a 1.5% (w/v) agarose gel to confirm the correct fragment size. PCR reactions were treated with *Dpn* I restriction enzyme to remove the original pDNA template and then heated to 85°C to denature the enzyme. Each overlap PCR reaction was diluted 1/100 and then combined to be used as template for a second round of high fidelity PCR using the vector specific primers (T7promoter and T7terminator). The two PCR products provided overlapping sequences of 34bp (which incorporated the D330A/E332A mutations) to allow annealing for PCR. An aliquot of the PCR reaction was visualised on a 1.5% (w/v) agarose gel (Figure 6.9) and then purified as before.



The mutated PCR amplicon was cloned in an identical manner to the mutated *T.in* Pol I fragment. Colonies were screened by PCR to confirm the presence of recombinant colonies. Five recombinant colonies were ExoSAP treated and DNA sequenced using the T7promoter primer to confirm the successful introduction of the D330A and E330A mutations.

### Kosmotoga oleania 3'-5' exo KO

Plasmid containing the *K.ol* *polA* gene was purified for use as the DNA template for a site-directed mutagenesis PCR, as reported previously, to generate the *T.in* and *T.at* *polA* mutants. Two overlapping PCR reactions were run to generate the D327A and E329A mutations. The first reaction contained primers T7promoter and Kol\_3-5exoKO\_L to give a 259bp product and the second Kol\_3-5exoKO\_U and T7terminator to give a 1868bp product. The mutagenic primer sequences are detailed in Figure 6.10a. Amplified DNA products were visualised on a 1.5% (w/v) agarose gel to confirm the correct fragment size. PCR reactions were treated with *Dpn* I restriction and then heated to 85°C to denature the enzyme. Each overlap PCR reaction was diluted 1/100 and then combined to be used as template for a second round of high fidelity PCR using the vector specific primers (T7promoter and T7terminator). The two PCR products provided overlapping sequences of 49bp (which incorporated the D327A/E329A mutations) to allow annealing for PCR. An aliquot of the PCR reaction was visualised on a 1.5% (w/v) agarose gel and then purified.



**Figure 6.10**

Site directed mutagenesis of *K.ol*\_D327A/E329A DNA pol I. (a) The mutagenic primers used to mutate the two key 3'-5' exonuclease residues in *K.ol* Pol I. M: 500µg Lambda *EcoR* I/ *Hind* III marker. (b) An agarose gel showing the 2078bp mutagenic *polA* PCR fragment.

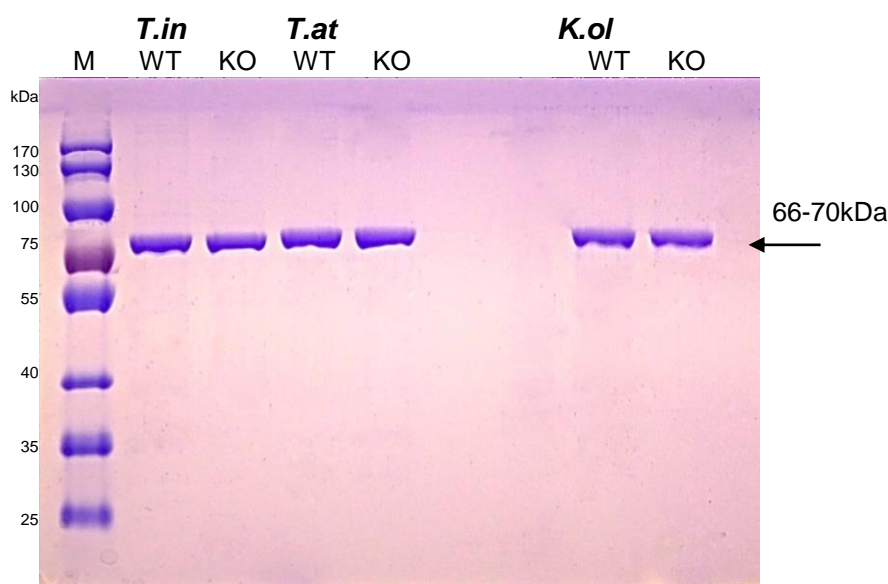
The mutated PCR amplicon was cloned in an identical manner to the mutated *T.in* and *T.at* Pol I fragments. Colonies were screened by PCR to confirm the presence of recombinant colonies. Five recombinant colonies were ExoSAP treated and DNA sequenced using the T7promoter primer to confirm the successful introduction of the D327A and E329A mutations.

## Expression and purification

The 3'-5' exonuclease Pol I mutants were re-labelled:

- *T.in*\_D330A/E332A Pol I was re-labelled ***T.in(exo-)* Pol I**
- *T.at*\_D330A/E332A Pol I was re-labelled ***T.at(exo-)* Pol I**
- *K.ol*\_D327A/E329A Pol I was re-labelled ***K.ol(exo-)* Pol I**

The clones containing *T.in(exo-)*, *T.at(exo-)* and *K.ol(exo-)* Pol I were over-expressed on a large scale and the enzymes were purified using the same three column chromatography approach, reported in Chapter 4. Each purified enzyme was quantified using the Qubit<sup>®</sup> protein concentration method and 1 µg was analysed by SDS-PAGE alongside the wild-type enzyme to confirm the purity of the sample (Figure 6.11).

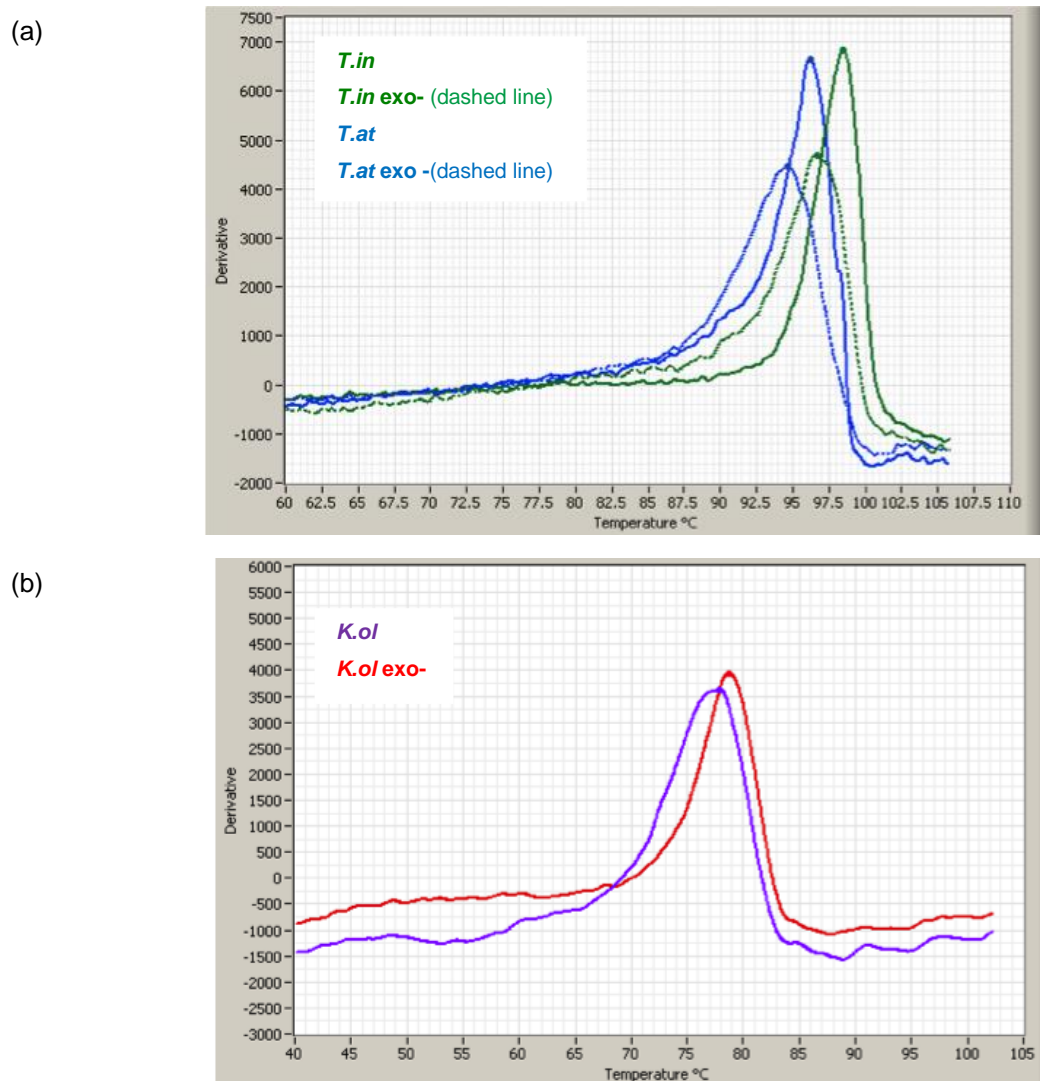


**Figure 6.11**

12% SDS-PAGE comparing 1 µg purified wild-type (WT) and 3'-5' exonuclease KO mutants (KO). The protein standard M: PageRuler. The overexpressed Pol I with expected  $M_w$  is identified by an arrow.

## Characterisation

The purified 3'-5' exonuclease (exo-) Pol I mutant enzymes were characterised using the same methods detailed in Chapter 5 and compared to the wild-type enzyme. 1 µg protein was first analysed in the SYPRO®-Orange thermal shift assay, to compare the thermostability of the mutant and WT enzymes. The three Pol I mutants each achieved a higher temperature of denaturation when compared to the WT enzyme (Figure 6.12).



**Figure 6.12**

Thermal shift assay comparing the denaturation profile of 1 µg (a) *T.in* and *T.at* Pol I vs *T.in*(exo-) and *T.at*(exo-) Pol I, (b) *K.ol* vs *K.ol*(exo-) Pol I, in 1X iBuffer, in the presence of 10X SYPRO®-Orange, following the method in Chapter 2.6.3.

The Pol I mutants were further characterised using the 5'-3' DNA polymerase primer extension assay, detailed in Chapter 2.6.10. Previously reported in Chapter 5.2.10, the assay enabled a comparison of the 5'-3' DNA polymerase activity of each WT and mutant enzyme under the same conditions. The effect of the mutations on the ability to incorporate nucleotides, to a pre-annealed ssDNA oligo, could be observed using this assay. Due to the potential for residual 3'-5' exonuclease activity remaining, the primer extension assays were reported using the 3'-phosphorothioated modified primers. This prevented degradation of the 3'-ends of the ssDNA oligonucleotides in the reaction.

The M13mp18 ssDNA primer extension assay reported identical activities of each mutant enzyme to its parent wild-type enzyme (Table 6.3), indicating the mutations had no effect on the 5'-3' DNA polymerase activity of the enzyme.

	Specific activity ( <i>Bst</i> -like units)
<i>Tin</i>	15000
<i>Tin</i> exo-	15000
<i>Tat</i>	18750
<i>Tat</i> exo -	18750
<i>Kol</i>	180000
<i>Kol</i> exo-	180000

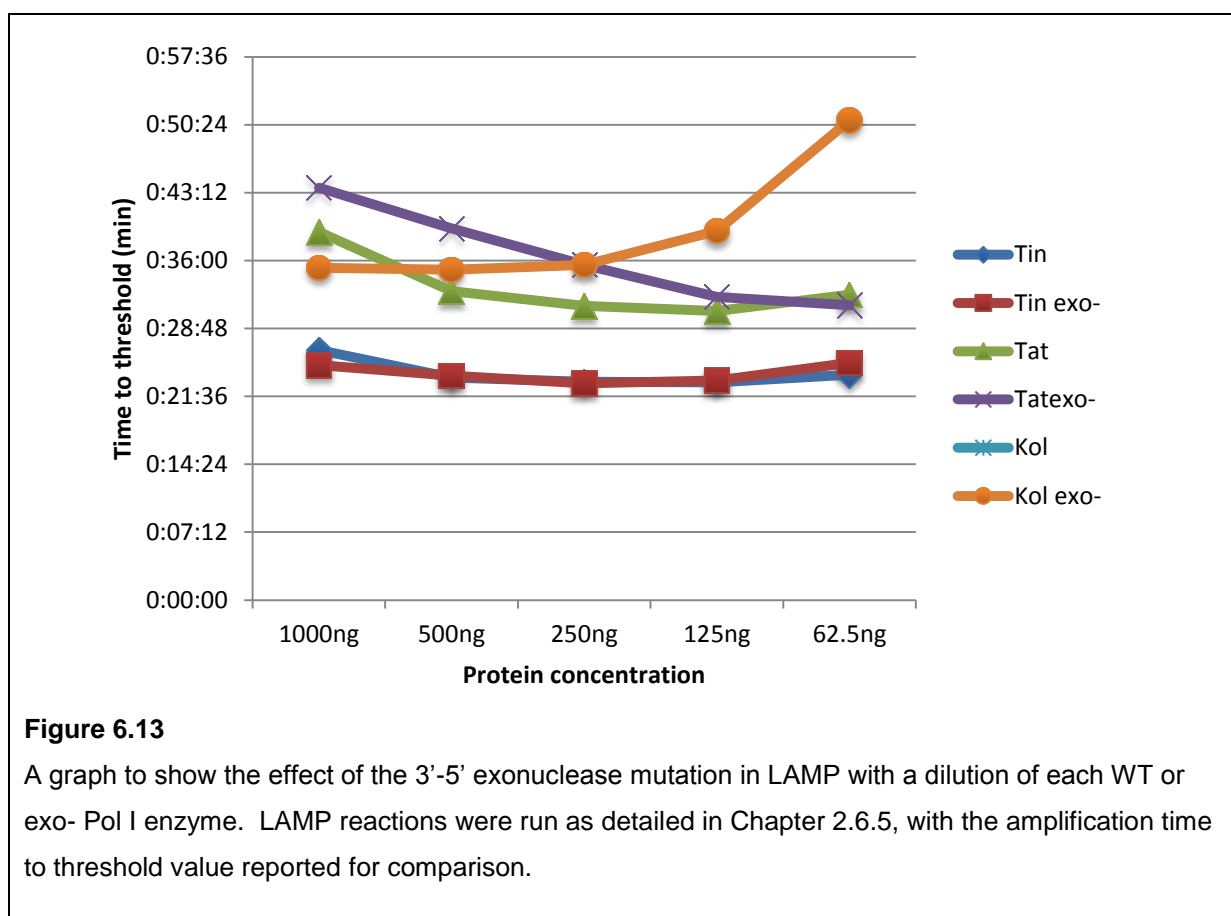
**Table 6.3**

A comparison of the extension rates of the WT and 3'-5' exo- mutant Pol I enzymes, generated using the M13mp18 ssDNA primer extension assay (method 2.6.10), with the specific activity calculated as *B.st* Pol I equivalent units/mg.

A dilution of each WT and exo- enzyme was used in a LAMP reaction to compare the effect on amplification speed. The results are summarised in Figure 6.13, which plots the amplification time to threshold value with the dilution of each enzyme used in the LAMP reaction. It was expected that the exonuclease mutant Pol I enzymes would show a reduced time to amplification, if the proposed 3'-5' exonuclease activity was indeed present in the wild-type enzyme, and was degrading the ssDNA LAMP primers as expected. *T.in*(exo-) Pol I (red plot) reported almost identical amplification times with the wild-type enzyme (blue plot) across all enzyme dilutions. *T.at*(exo-) Pol I (purple plot) again shows close amplification times to the wild-type enzyme (green plot). These result suggests the exo-mutation had no effect on enhancing the LAMP reaction. This may because the predicted



exonuclease activity of the wild-type enzyme is either not present, or is too weak to be detrimental to the LAMP reaction.

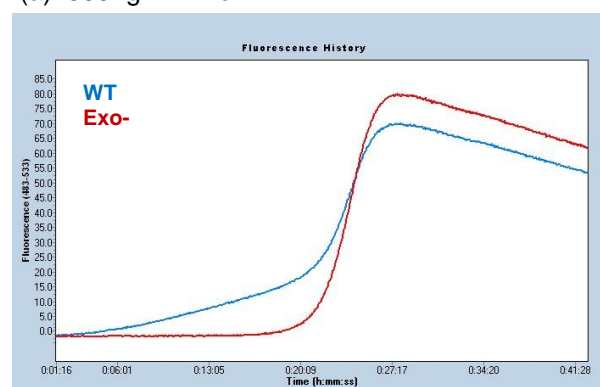


Reviewing the real-time LAMP amplification plots in Figure 6.14, however, clearly demonstrates the positive effect of the 3'-5' exonuclease mutation to the LAMP reaction. The activity of the 3'-5' exonuclease domain can be seen to have been successfully reduced for each Pol I, preventing primer degradation that was believed to be responsible for the increase in fluorescence over background prior to amplification.

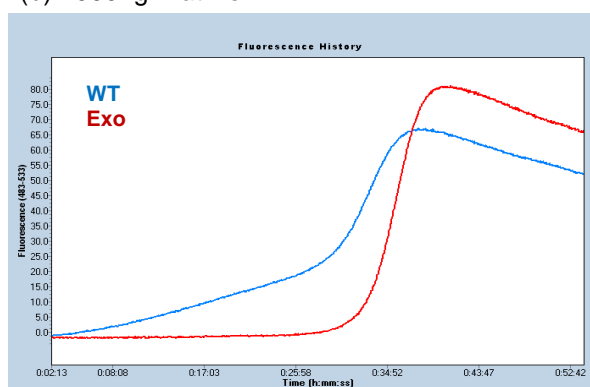
Figure 6.14 shows the real-time LAMP amplification of the WT (blue) and exo- (red) in LAMP. 500ng of *T.in*(exo-) Pol I is required to demonstrate the effect of primer degradation in a LAMP reaction, shown by the blue line increasing the fluorescence over background before amplification in Figure 6.14a. A similar result can be seen for *T.at*(exo-) Pol I, in Figure 6.14b. Here, 1000ng enzyme stock is required to show the effect of the 3'-5' exo mutations. The high level of protein required to demonstrate 3'-5' exonuclease activity indicates the activity is present, but relatively weak. This is in contrast to *K.ol* Pol I, which requires the 3'-5' exonuclease mutation to successfully work in LAMP (Figure 6.14c). This

result implies the wild-type *K.o/* enzyme degrades the LAMP reaction primers sufficiently to prevent LAMP occurring, indicating significantly stronger 3'-5' exonuclease activity compared to *T.in* and *T.at* Pol I enzymes. *K.o/*(exo-) Pol I amplification still shows an increase in fluorescence over background prior to LAMP amplification, suggesting the exonuclease activity has not been fully removed from the enzyme.

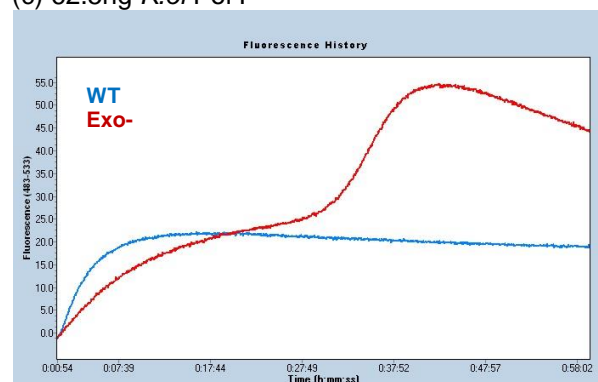
(a) 500ng *T.in* Pol I



(b) 1000ng *T.at* Pol I



(c) 62.5ng *K.o/* Pol I



**Figure 6.14**

Real-time LAMP amplification data for the WT (blue) and exo- (red) Pol I enzymes. LAMP reactions were run as detailed in Chapter 2.6.5, with the amplification time to threshold value reported in Figure 6.13. (a) 500ng (64u *B.st* Pol I equivalent) *T.in* Pol I (b) 1000ng (128u *B.st* Pol I equivalent) *T.at* Pol I, (c) 62.5ng (8u *B.st* Pol I equivalent) *K.o/* Pol I.

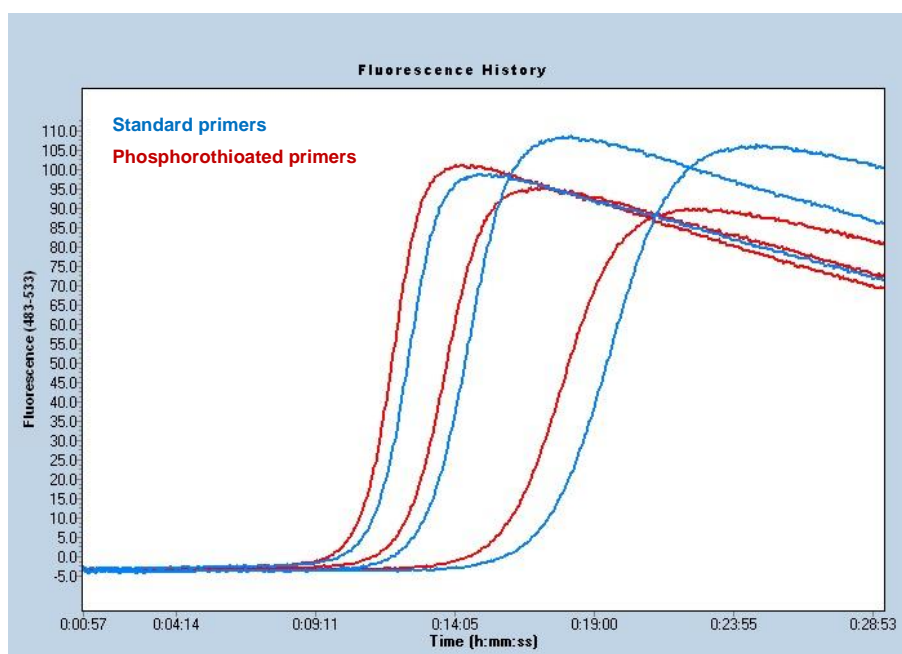
The action of the 3'-5' exonuclease on the LAMP primers is expected to have significantly affected the ability of the enzymes to function in LAMP. For this reason, an identical set of *A.pe* LAMP primers were designed to incorporate a 3'-phosphorothioated modification at the n-1 position at the 3' end of the six LAMP oligos (Figure 6.15). This modification prevents the 3'-5' exonuclease domain removing bases from the single-stranded oligonucleotides, thus leaving all the primers available to the enzyme in the reaction.

FIP(F1c+F2)	CGTACAGCCTGAGATCGACAAGGTAGTAATGCTAGAGCATCT <sup>G<sup>s</sup></sup> G
BIP(B1c+B2)	GTGTAGACGAAGAGGCGGTGCTCAACACTACAGCCA <sup>C<sup>s</sup></sup> G
F3	GGCAATAGTACATGAAGG <sup>C<sup>s</sup></sup> T
B3	CATAAGATTGTCCACGGCT <sup>A<sup>s</sup></sup> T
LoopF	AGCGTGACCTACAGCA <sup>A<sup>s</sup></sup> C
LoopB	AAGGATGAGACGAGAAGTGT <sup>T<sup>s</sup></sup> G

**Figure 6.15**

Alternative *A.pe* LAMP primer set incorporating a modified base at n-1 (Yellow).

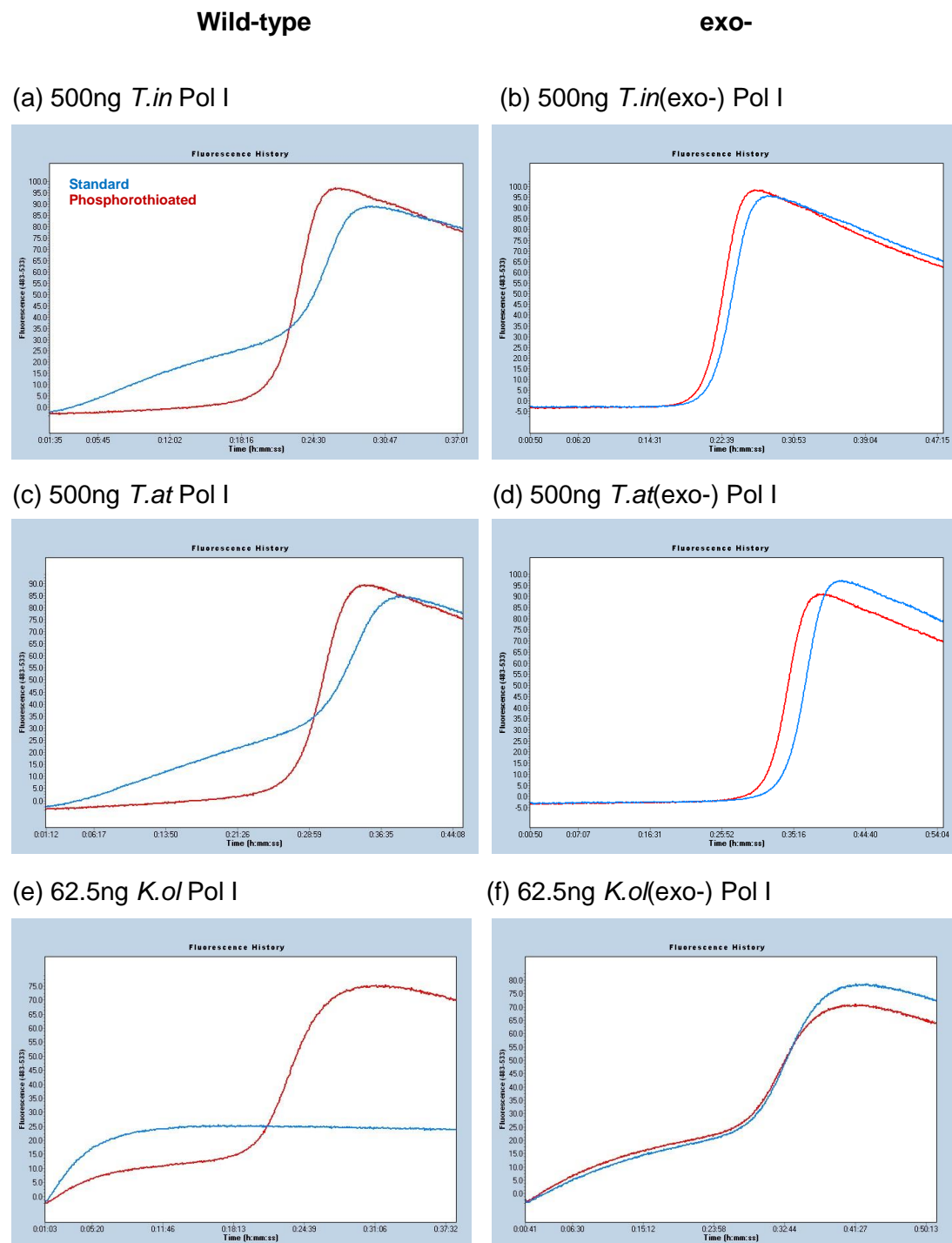
The effect of the 3'-phosphorothioate modification on amplification times was first investigated using a dilution of the 3'-5' exonuclease deficient, *G.me*\* Pol I (Figure 6.16). LAMP amplification times were shown to be comparable, with the insignificant variation in amplification time attributed to the oligonucleotide preparation from the supplier rather than the ability of the Polymerase to incorporate the primer modification. Inconsistent batch to batch primer variations have been identified at GeneSys Biotech Ltd., especially when varying the scale of oligonucleotide production. Modified primers can only be ordered on a larger scale.



**Figure 6.16**

Effect of phosphorothioated primers (Red) vs standard LAMP primers (Blue). LAMP reactions were run as detailed in Chapter 2.6.5, *G.me*\* Pol I, dilutions: 62, 125, 250ng. The time to amplification is observed to be reduced with increasing Pol I enzyme in the reaction.

The amplification showed the modified primers did not significantly affect the amplification of the LAMP products. LAMP reactions were therefore repeated to compare wild-type and 3'-5' exonuclease minus mutants in the presence of either standard or modified *A.pe* LAMP primer sets (Figure 6.17).



**Figure 6.17**

LAMP reactions using either phosphorothioate modified LAMP primers (red plot) or standard LAMP primers (blue plot). The significant result from this assay can be observed in (e) where the WT *K.o* Pol I exclusively requires modified primers for a successful LAMP reaction.

The modified LAMP primers show the same effect on a LAMP reaction as the 3'-5' exonuclease mutation, i.e. both prevent primer degradation and reduce the increase in background fluorescence prior to amplification. Again, wild-type *K.o*/ Pol I can be seen to degrade the un-modified primers (Figure 6.17e), preventing LAMP amplification. This confirms the results reported are due to the 3'-5' exonuclease activity of the enzymes.

The increase in background fluorescence observed with *K.o*/(exo-) Pol I, in the presence of modified primers (Figure 6.17e, red plot), confirmed that the fluorescence is not just occurring due to primer degradation. The phosphorothioated primers cannot be digested, and therefore the increase is not due to residual 3'-5' exonuclease. In the wild-type 3'-5' exonuclease active reactions, with standard primers, the increase in background is thought to be due to the annealing of degraded products, resulting in linear template extension, gradually increasing dsDNA in the reaction, available for the fluorescent intercalating dye to bind. The *T.in* and *T.at* wild-type Pol I are reasoned to have significantly weaker 3'-5' exonuclease activity. This is because a low level of primer degradation leads to a slight increase in background fluorescence with a significant quantity of enzyme (Figure 6.17a/c – blue) but the exponential increase in LAMP product, generated by the LAMP primers still available, is faster than the background, as indicated by a positive amplification curve (Figure 6.17a/c – red).

The observation of *K.o*/(exo-) Pol I still demonstrating this increase in background fluorescence is noted to be enzyme specific. It is not observed with the other polymerases, wild-type or 3'-5' exonuclease deficient/reduced. To investigate this result further, the LAMP reactions were repeated without the DNA template and or reaction primers. The increase in fluorescence was only removed when the primers were omitted from the reaction. It is therefore reasoned that the *K.o*/ Pol I is further interacting with the primers, possibly leading to the formation of dsDNA primer dimers.

*T.aq* Pol I and *T.ma* Pol I are enzymes of great commercial value in the biotechnology industry due to their high thermostability enabling their use in PCR amplification methods. *T.aq* and *T.ma* Pol I have not previously been reported to show strand-displacement activity and the results in Chapter 5 confirmed this. Singh *et al.* (2007) have proposed three key residues that may be responsible for strand-displacement activity in *E.co* DNA Pol I. They reported Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup> are a pre-requisite for strand separation and they showed each position, when mutated to alanine, removed the ability of the Pol I to strand displace.

**Figure 6.18**  
ClustalW sequence alignment highlighting the Singh *et al.* (2007) proposed 3 key strand-displacement residues, highlighted in yellow. The local DNA polymerase motifs are highlighted in grey. Thr<sup>769</sup> of *T.ma* DNA Pol III identified in pink.

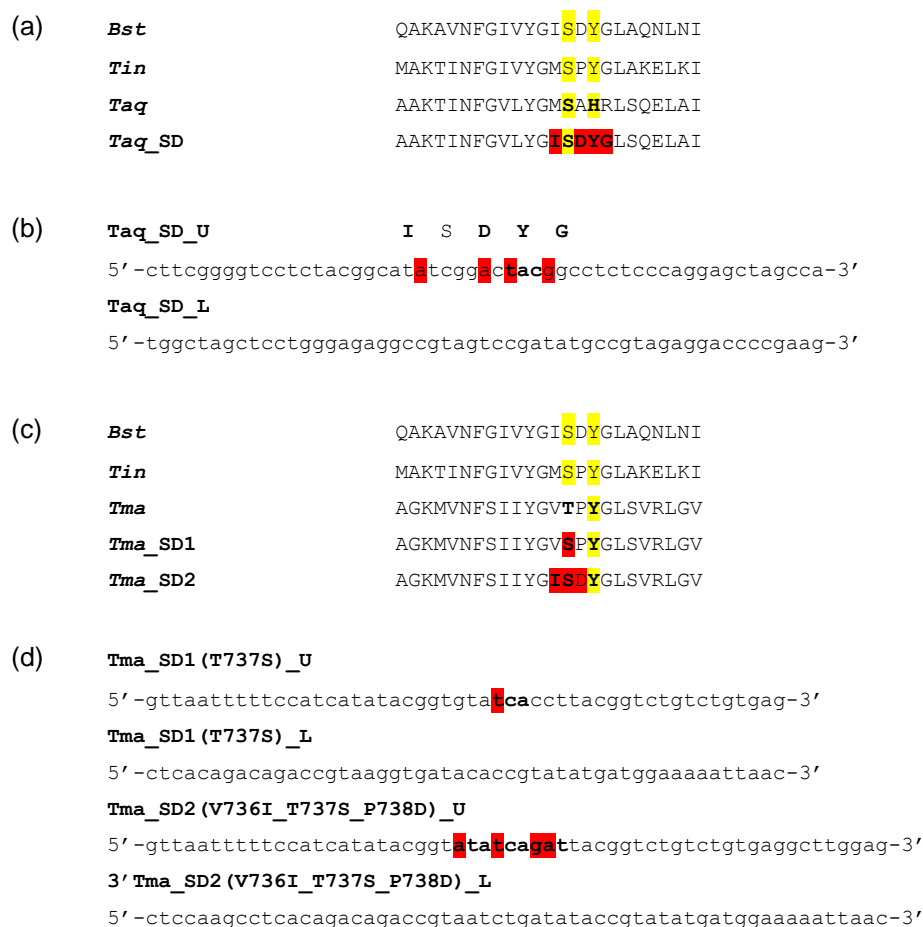
Sequence alignments show the three residues to be highly conserved among known strand-displacing polymerases (Figure 6.18). However:

- *T.ma* Pol I has the required Phe<sup>771</sup> and Arg<sup>841</sup> residues but lacks the Serine, where instead there is Thr<sup>769</sup>.
- *T.aq* Pol I appears to qualify as a strand-displacing enzyme under the rules of residue requirements (Singh *et al.*, 2007). *T.aq* Pol I contains Ser<sup>769</sup>, Arg<sup>871</sup> but has a His<sup>771</sup>, an aromatic residue that is suggested to still enable activity by the authors.



## Cloning expression and purification

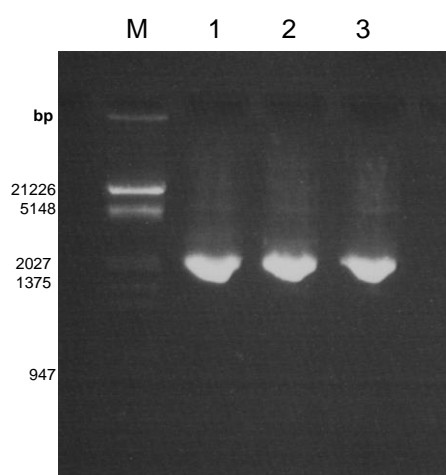
The *T.aq* and *T.ma* enzymes were mutated to *B.st* Pol I-like residues to incorporate sequence known to provide strong strand-displacement activity following the mutations highlighted in Figure 6.19. Two mutants were designed for *T.ma* Pol I (i) a single point mutation to mutate Thr<sup>769</sup> to Ser<sup>769</sup>, and (ii) multiple residues to mutate the local region to a *B.st* Pol I-like enzyme.



**Figure 6.19**

(a) The amino acid residues targeted for mutagenesis (red) to create the mutant *T.aq*\_SD. (b) the mutagenic primer sequences to create *T.aq*\_SD. (c) The amino acid residues targeted for mutagenesis (red) to create the mutant *T.ma*\_SD1 and *T.ma*\_SD2. (d) the mutagenic primer sequences to create *T.ma*\_SD1 and *T.ma*\_SD2.

Plasmid DNA, containing either the *T.aq* or *T.ma* DNA *polA* gene was purified for use as the DNA template for a site-directed mutagenesis PCR. Two overlapping PCR reactions were used to generate each mutation, using the mutagenic primers detailed in Figure 6.19. Mutant *T.aq*\_SD1 was created using primers T7promoter and Taq\_SD\_L to give a 1351bp product; and Taq\_SD\_U and T7terminator to give a 637bp product. Mutant *T.ma*\_SD1 was created using primers T7promoter and Tma\_SD1\_L to give a 1474bp product; and Tma\_SD1\_U and T7terminator to give a 636bp product. Finally, mutant *T.ma*\_SD2 was created using primers T7promoter and Tma\_SD2\_L to give a 1525bp product: and Tma\_SD2\_U and T7terminator, to give a 636bp product. Each PCR reaction was treated with *Dpn* I restriction enzyme to digest the original template pDNA and then heated to 85°C to denature the enzyme. Each overlap PCR reaction was diluted 1/100 and then combined to be used as template for a second round of high fidelity PCR using the vector specific primers (T7promoter and T7terminator). An aliquot of the PCR reaction was visualised on a 1.5% (w/v) agarose gel (Figure 6.20) and then purified using a spin column purification kit.

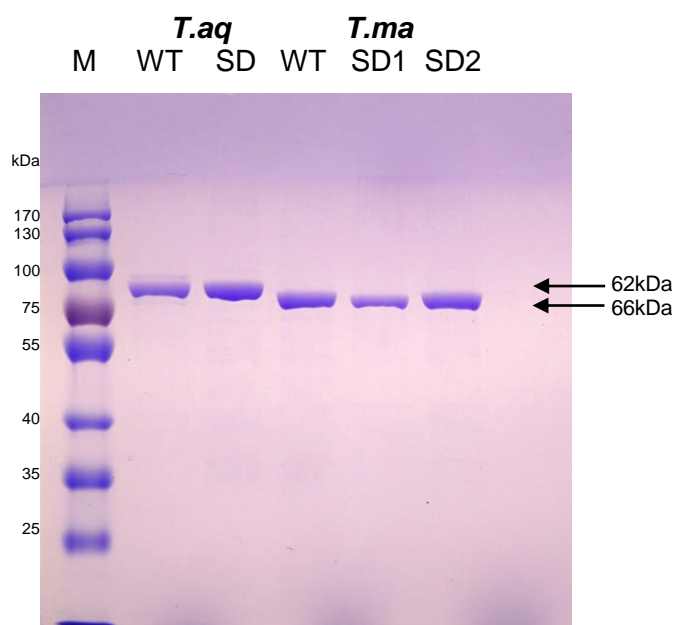


**Figure 6.20**

An agarose gel showing the mutagenic *polA* PCR fragments generated using T7primers. Site directed mutagenesis of 1: *T.aq*\_SD1 (1910bp fragment) , 2: *T.ma*\_SD1 (2081bp fragment), 3: *T.ma*\_SD2 DNA Pol I (2081bp fragment). M: 500 µg Lambda DNA *EcoR* I / *Hind* III marker.

The mutated PCR amplicon was *Nde* I and *Bam*H I digested to allow directional cloning into the pET24a+HIS vector. The ligated clone was transformed into *E.coli* KRX(pRARE2). Five recombinant colonies were each screened by PCR using the vector specific T7 primers, and the amplicons were ExoSAP treated and DNA sequenced using the T7promoter primer to confirm the successful introduction of the required SD mutations.

*T.aq*\_SD, *T.ma*\_SD1 and *T.ma*\_SD2 Pol I were over-expressed on a large scale and the recombinant proteins were purified using the three column chromatography approach, previously reported in Chapter 4.3. Purified enzyme was quantified using the Qubit® protein concentration method and 1 µg was electrophoresed on a 12% SDS-PAGE gel alongside the wild-type enzyme to assess the purity of the sample (Figure 6.21).

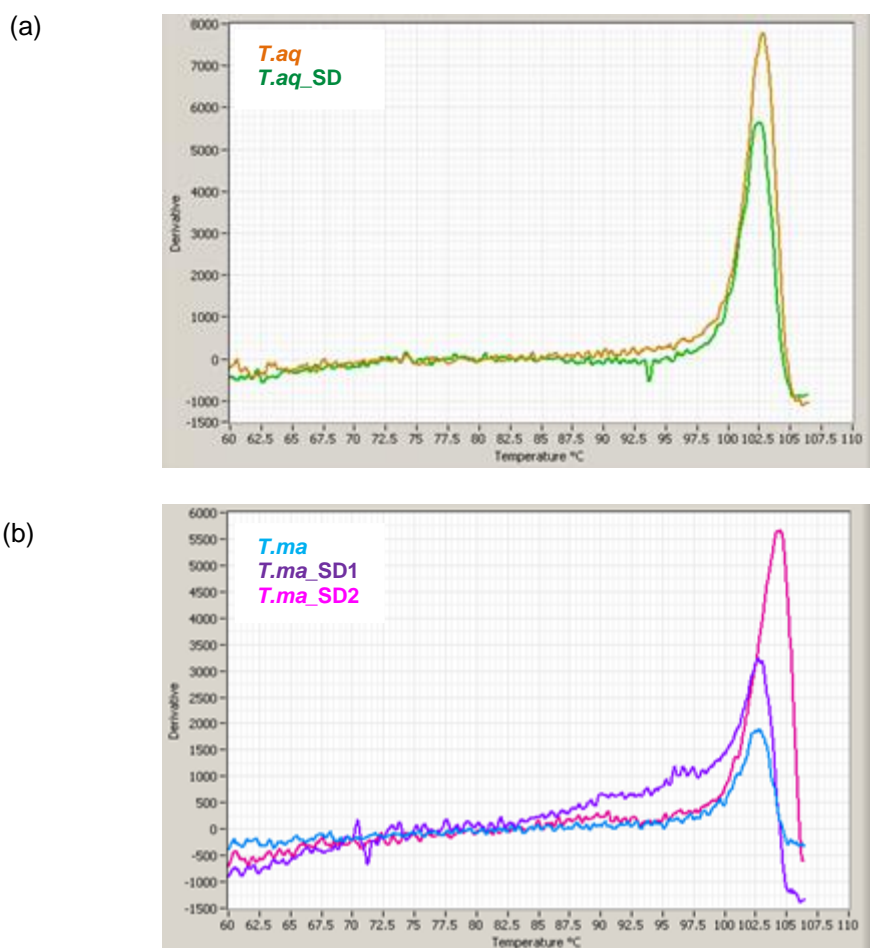


**Figure 6.21**

12% SDS-PAGE comparing 1µg purified wild-type (WT) and SD mutants (SD). The protein standard M: PageRuler. The overexpressed Pol I with expected  $M_w$  is identified by an arrow. It should be noted that the expected  $M_w$  of the protein and its migration pattern through the gel is effected by the pI of the protein (detailed in Table 4.1).

## Characterisation

Purified Pol I mutant enzymes were characterised using the methods detailed in Chapter 5 and compared to the wild-type enzyme. 1µg protein was analysed in the SYPRO®-orange thermal shift assay, to compare the thermostability of the mutant and WT enzymes. *T.aq*\_SD Pol I demonstrated a similar denaturation profile to the wild-type enzyme, reporting an identical  $T_m$  value (Figure 6.22b). *T.ma*\_SD1 reported an identical  $T_m$  value to the wild-type, but *T.ma*\_SD2 was shown to be marginally more thermostable, with a  $T_m$  increased by 1.4°C over *T.ma* Pol I



**Figure 6.22**

Thermal shift assay, using 1µg protein, comparing the denaturation profile of (a) *T.aq* (yellow plot) vs *T.aq*\_SD (green plot) Pol I, (b) *T.ma* (light blue plot) vs *T.ma*\_SD1 (purple plot) and *T.ma*\_SD2 (pink plot) Pol I, in 1X iBuffer, in the presence of 10X SYPRO®-Orange, following the method in Chapter 2.3.6.

To test the ability of the mutants to strand-displace DNA, a dilution of enzyme was run in a LAMP reaction, detailed in Chapter 2.6.5. No amplification was observed from any sample, indicating the *B.st* Pol I-like residues did not confer the ability to strand-displace DNA, to the mutant enzymes.

It was necessary to confirm that the mutant enzymes were active, and that the strand-displacement result was not due to the absence of 5'-3' DNA polymerase activity. The M13mp18 ssDNA primer extension assay demonstrated that all mutants successfully extended the primer. The mutation of *T.aq* Pol I was shown to reduce activity further, to 225 units/mg. The *T.ma* Pol I mutants, however, showed increased primer extension activity, indicating the *Bst*-like residues enhanced 5'-3' DNA polymerase activity (Table 6.4).

	Specific activity ( <i>Bst</i> units/mg)
<i>Taq</i>	1000
<i>Taq_SD</i>	225
<i>Tma</i>	600
<i>Tma_SD1</i>	1400
<i>Tma_SD2</i>	900

**Table 6.4**

Extension rates generated using the M13mp18 ssDNA primer extension assay and calculated as *B.st* Pol I-equivalent units/mg.

Reporting activity in the 5'-3' DNA polymerase activity assay, confirmed that the strand-displacement mutations were not successful in enhancing either *T.aq*, or *T.ma* Pol I for use in LAMP.

## 6.2.4 Strand displacement 'SYR' motif

Singh *et al.* (2007) reported that three key residues (S, Y and R) were a pre-determinant for strand-displacement activity in *E.coli* DNA Pol I. *G.me*\* Pol I, a sequence and functional equivalent of *B.st* Pol I, and *T.in*(exo-) Pol I were chosen for site-saturation mutagenesis at the three key residues (Figure 6.23).

		SD1SD2	
<b>Tin</b>	MAKTINFGIVYGM	SPYGLAKELKIGRREAKAFIERYFERYPGVKRYMEQIVAEAREKGYV	789
<b>Gme*</b>	QAKAVNFGIVYGI	SDYGLAQNLNITRKEAAEFIERYFASFPGVKQYMENIVQEAKQKGYV	763
		SD3	
<b>Tin</b>	ETLFGRKRPLPDINSPNRTAREFAER	TAINTPIQGTAAADI IKLAMIKIHRIFKEKGFGR	849
<b>Gme*</b>	TLLHRRRYLPDITSRNFNVSFAER	TAMNTPIQGSAAADI IKKAMIDLAARLKEEQLQAR	823

**Figure 6.23**

Sequence locations of the three key strand-displacement residues reported by Singh *et al.* (2007). *G.me*\* Pol I SD1 (S717), SD2 (Y719), SD3 (R789). *T.in* DNA Pol I SD1 (S743), SD2 (Y745), SD3 (R815)

Small-intelligent primer mixes (Tang *et al.*, 2012), detailed in Appendix i.vi were designed for each selected amino acid position for targeting by site-saturation mutagenesis. Small-intelligent primer mix *GspM\_S717X\_U* was run with the T7terminator primer and small-intelligent primer mix *GspM\_S717X\_L* was run with the T7promoter primer. Amplified DNA products were visualised on a 1.5% (w/v) agarose gel to confirm the correct fragment size. PCR reactions were treated with *Dpn* I restriction enzyme to remove the original template pDNA and then heated to 85°C to denature the enzyme. Each overlap PCR reaction was diluted 1/100 and then combined to be used as template for a second round of high fidelity PCR using the vector specific primers (T7promoter and T7terminator). An aliquot of the PCR reaction was visualised on a 1.5% (w/v) agarose gel and then purified using a spin column purification kit.

Each additional site-saturation mutation was created in the same way, using the specific primer combinations reported in Table 6.5 and 6.6.

	<b><u>GspM</u> S717X</b>	<b><u>GspM</u> Y719X</b>	<b><u>GspM</u> R789X</b>
	5' – acttcggtatcggtttacggaatt <b>XX</b> <b>X</b> gattacggattggcgcaaaactt- 3'	5' – gtatcggtttacggaattagcgat <b>XXX</b> gg attggcgcaaaacttgaacat-3'	5' – tcaacgtccgcagttttgcagag <b>XXX</b> ac ggccatgaacacgccaattca-3'
12	GspM_S717X_NDT_U GspM_S717X_NDT_L	GspM_Y719X_NDT_U GspM_Y719X_NDT_L	GspM_R789X_NDT_U GspM_R789X_NDT_L
6	GspM_S717X_VMA_U GspM_S717X_VMA_L	GspM_Y719X_VMA_U GspM_Y719X_VMA_L	GspM_R789X_VMA_U GspM_R789X_VMA_L
1	GspM_S717X_ATG_U GspM_S717X_ATG_L	GspM_Y719X_ATG_U GspM_Y719X_ATG_L	GspM_R789X_ATG_U GspM_R789X_ATG_L
1	GspM_S717X_TGG_U GspM_S717X_TGG_L	GspM_Y719X_TGG_U GspM_Y719X_TGG_L	GspM_R789X_TGG_U GspM_R789X_TGG_L

**Table 6.5**

*G.me*\* Pol I small-intelligent primer mix for the three strand-displacement residues targeted for site-saturation mutagenesis.

	<b><u>T.in(exo-)</u> S743X</b>	<b><u>T.in(exo-)</u> Y745X</b>	<b><u>T.in(exo-)</u> R815X</b>
	5' – actttggcattgtttacggcatg <b>XX</b> <b>X</b> cctt <b>ac</b> gggtctggcgaaagaact- 3'	5' – gcattgtttacggcatgagtcct <b>XXX</b> gg tctggcgaaagaactcaaaattg-3'	5' – gtacggcgcgagtttgccgag <b>XXX</b> ac ggctataaacactcctattca-3'
12	Tin_S743X_NDT_U Tin_S743X_NDT_L	Tin_Y745X_NDT_U Tin_Y745X_NDT_L	Tin_R815X_NDT_U Tin_R815X_NDT_L
6	Tin_S743X_VMA_U Tin_S743X_VMA_L	Tin_Y745X_VMA_U Tin_Y745X_VMA_L	Tin_R815X_VMA_U Tin_R815X_VMA_L
1	Tin_S743X_ATG_U Tin_S743X_ATG_L	Tin_Y745X_ATG_U Tin_Y745X_ATG_L	Tin_R815X_ATG_U Tin_R815X_ATG_L
1	Tin_S743X_TGG_U Tin_S743X_TGG_L	Tin_Y745X_TGG_U Tin_Y745X_TGG_L	Tin_R815X_TGG_U Tin_R815X_TGG_L

**Table 6.6**

*T.in(exo-)* Pol I small-intelligent primer mix for the three strand displacement residues targeted for site-saturation mutagenesis.

The mutated pool of PCR amplicons were *Nde* I and *Bam*H I digested to allow directional cloning into the pET24a+HIS vector. The ligated clones was transformed into *E.coli* KRX(pRARE2) and colonies screened to confirm the presence of recombinant colonies. Fifty recombinant colonies were screened by PCR using the vector specific T7 primers and the amplicons, ExoSAP treated and DNA sequenced using the T7terminator primer to confirm the presence and identity of each mutation. Two rounds of screening were required to identify the 19 variations of amino acids at each position.

With twenty different clones, for each of the three residues, for two different polymerases (120 clones), 100ml large scale expression and purification was not feasible in this study. 2ml 96 deep-well plates were therefore employed for 1ml TB culturing of the individual mutants. A single colony of each clone was inoculated into 3mls LB (kan) and grown overnight at 37°C (275rpm). 10µl of this starter culture was transferred to 1ml sterile TB (kan) pre-aliquoted into a 2ml 96 deep-well plate and incubated at 37°C (275rpm) for 4h at a 45°C angle to provide sufficient aeration to the media. At an OD<sub>600</sub> of 0.8 the temperature of the culture was reduced to 24°C to aid solubility and correct protein folding within the expression strain. Once the OD<sub>600</sub> had reached 1.0, the cultures were induced with an IPTG/L-rhamnose mix to provide a final concentration of 1mM and 0.1% respectively. The cultures were incubated for a further 22h to an OD<sub>600</sub> of 8.5. The plates were centrifuged to pellet the cells and the supernatant was removed. Pellets were resuspended in 1ml Lysis buffer (see 2.1.2) and sonicated for 10 sec *in situ*. The plates were then centrifuged as before to remove unwanted cell debris, leaving the soluble protein in the supernatant. A sample of each supernatant was electrophoresed on a 12% SDS-PAGE gel for confirmation of expression. The protein gels confirmed each mutant had expressed with the same protein yield allowing their direct comparison.

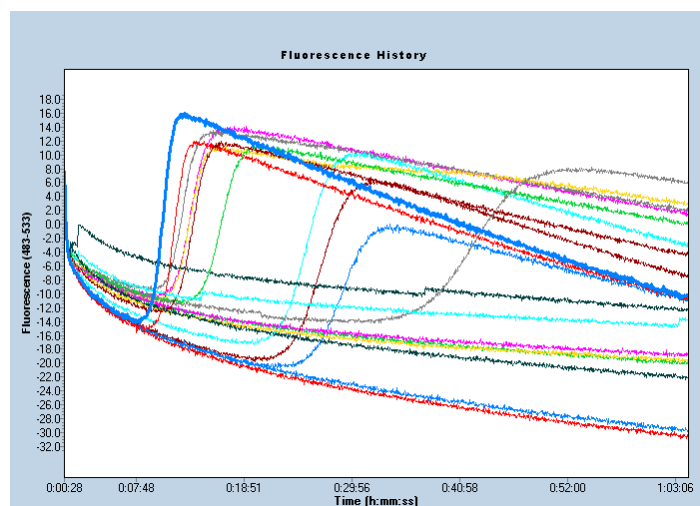


### Characterisation

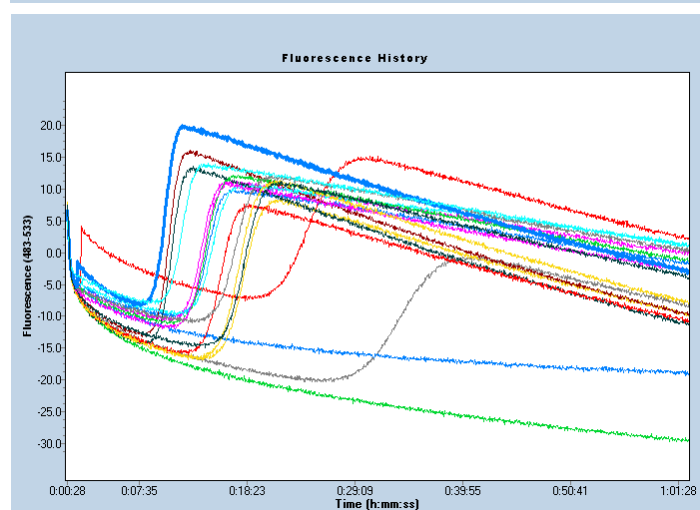
Due to the large number of clones, purification was not possible. Mutants were incubated at 65°C to remove the thermo-labile *E.coli* host protein, including *E.coli* DNA Pol I, reported to denature below 50°C (Karantzeni et al., 2003). The supernatant was further diluted for comparison in LAMP. 2µl of a 1/10 dilution of supernatant was added directly to a 25µl LAMP reaction pre-aliquoted in a 96-well PCR plate and run at 65°C on a LightCycler®-480 for real-time amplification. This dilution was expected to provide a reduction in contaminants from the crude extract, and reduce the expressed Pol I protein to suitable levels for comparison. A dilution of purified enzyme of known activity was run as a control to estimate the level of protein present within the samples.

The LAMP reaction trace files, for each mutation, can be seen in Figure 6.25 and 6.26 below. Positive amplification of the *A.pe* gDNA target is clearly visible from the crude extract wild-type enzymes samples (Bold blue plot in Figure 6.25, and bold red line in Figure 6.26). This result confirms the un-purified Pol I, in the presence of the crude extract, is sufficient to report LAMP activity.

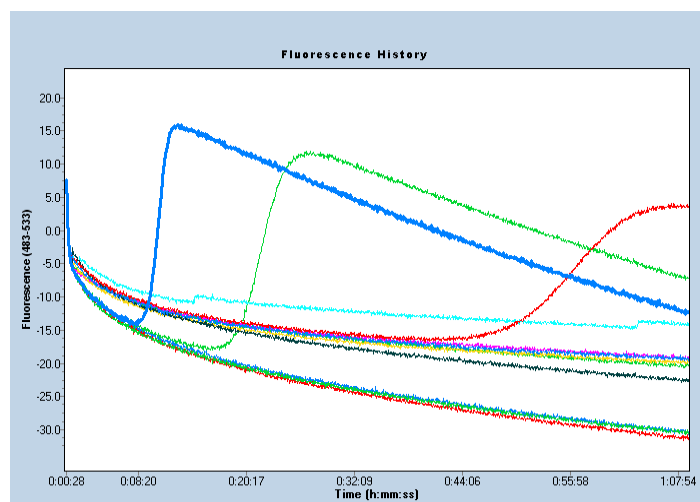
(a) Ser717X



(b) Tyr719X



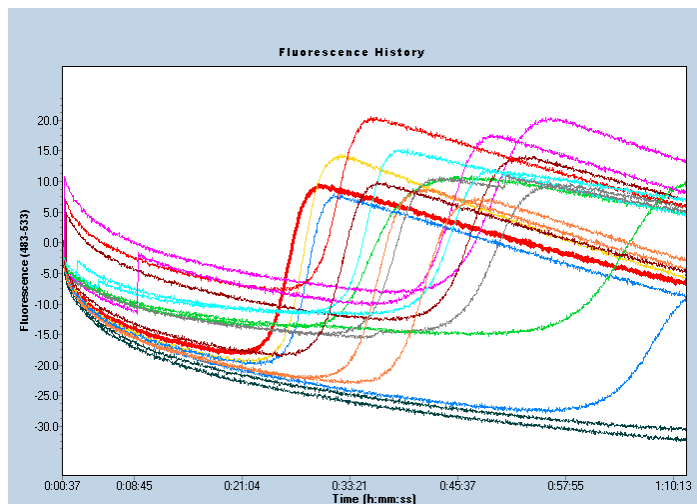
(c) Arg789X



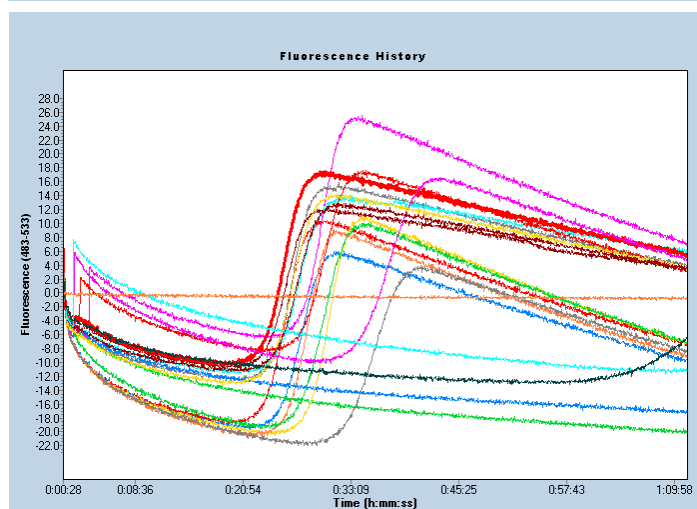
**Figure 6.25**

*G.me*\* Pol I mutants in LAMP: LAMP reactions were run as detailed in Chapter 2.6.5. *G.me*\* Pol I (WT) (Blue plot - BOLD), (a) 9 active S717X mutants (L,M,T,Q,C,G,R,H,D). (b) 16 active Y719X mutants (V,I,L,M,F,Y,W,S,T,N,C,G,R,H,K,E). (c) 2 active R789X mutants (KH). Remaining mutants and the un-induced control are shown as flat lined, non-amplified results.

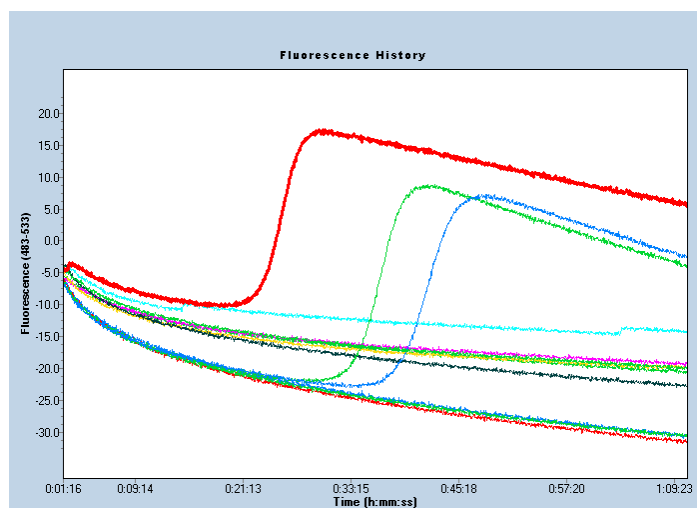
(a) Ser743X



(b) Tyr745X



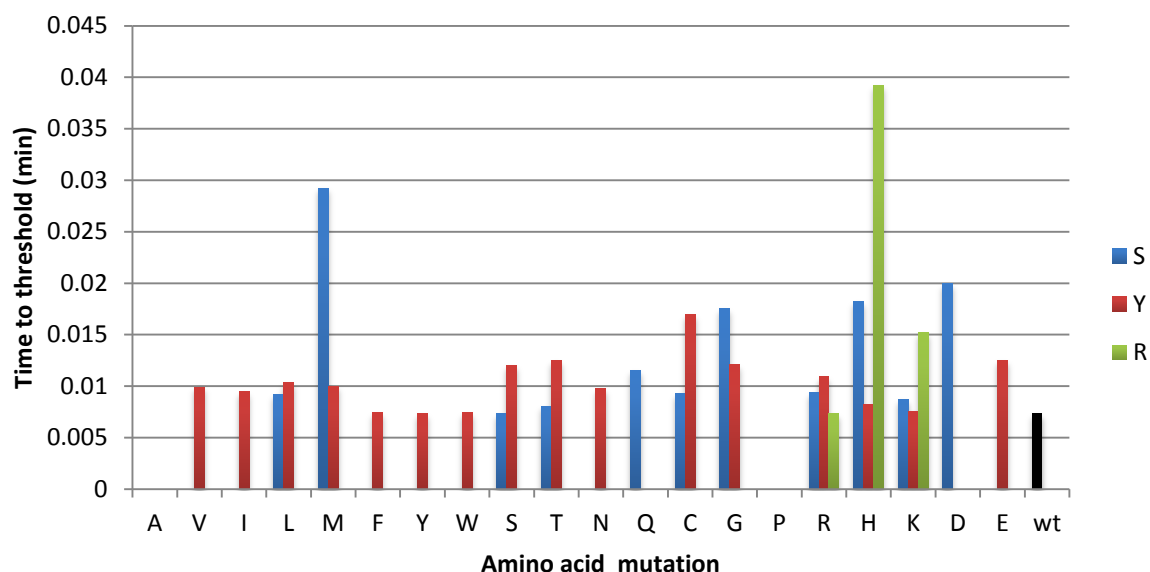
(c) Arg815X



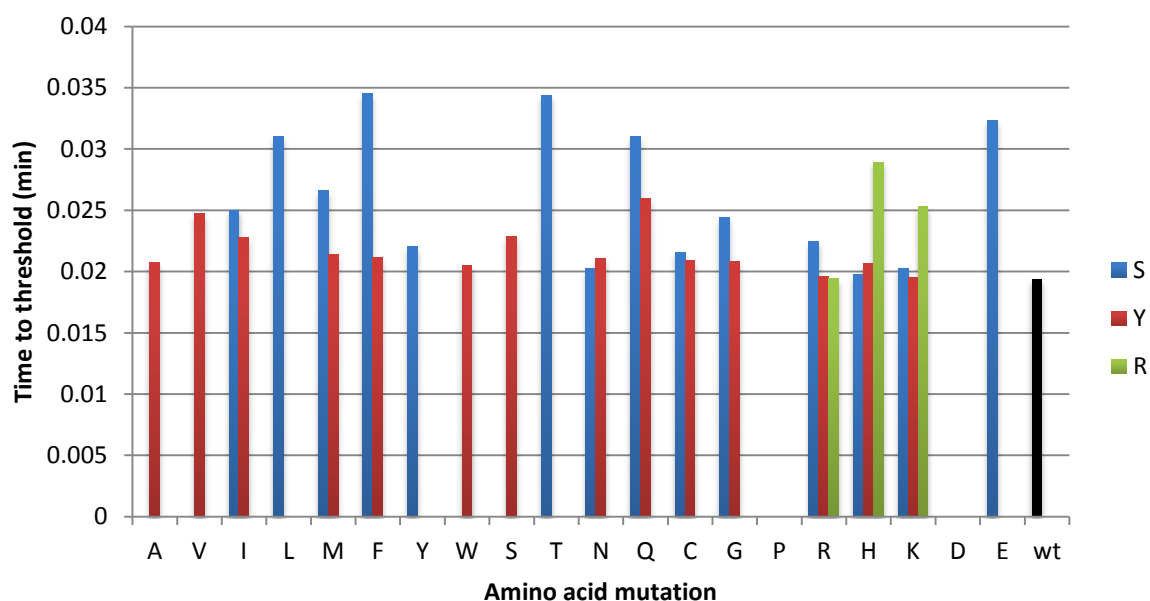
**Figure 6.26**

*T.in(exo-)* Pol I mutants in LAMP: LAMP reactions were run as detailed in Chapter 2.6.5. *Tin(exo-)* Pol I (WT) (red plot - BOLD), (a) 9 active S743X mutants (LMTQCGRHD). (b) 16 active Y745X mutants (V,I,L,M,F,Y,W,S,T,N,C,G,R,H,K,E). (c) 2 active R815X mutants (KH). Remaining mutants and the un-induced control are shown as flat lined, non-amplified results.

(a) *G.me\** DNA Pol



(b) *T.in(exo-)* DNA Pol



**Figure 6.27**

LAMP amplification time to threshold values reported from the data acquired in Figure 6.25 and 6.26 (a) *G.me\**\_SYR Pol I mutants. (b) *T.in(exo-)*\_SYR Pol I mutants. Bars indicate successful amplification by each mutated enzyme, identifying its ability to strand-displace. A blue bar represents a mutation of the key serine residue, a red bar represents a mutation of the key tyrosine residue, and a green bar represents a mutation of the key arginine residue. Black bars represent the wild-type (wt) enzyme, containing each of the SYR residues).

The LAMP amplification time to threshold values were taken from each positive reaction and further displayed in Figures 6.30 and 6.31. A large number of mutant Pol I enzymes were shown to be active in LAMP, demonstrated by the increase in fluorescence over background in Figures 6.25a/b/c and 6.26a/b/c. *G.me*\* and *T.in* Pol I both permit a wide variety of mutations in place of the serine and tyrosine position. However, the key arginine residue can be observed to absolutely require a positively charged residue in that position to facilitate LAMP.

Site-saturation mutagenesis of the 3 key SYR residues using *G.me*\* Pol I clearly shows several mutations to yield active strand-displacing polymerases. This is in direct contrast to the Singh *et al.* (2007) paper, based on *E.coli* Klenow, that indicated the three residues were each essential for activity. Figure 6.27a highlights the tyrosine residue (red bar) to be active with all but 4 of the possible amino acid mutations. The serine residue (blue bar) is also well tolerated with 12 of the possible mutations active. The arginine residue (green bar) only shows activity with a positive amino acid residue in that position.

Site-saturation mutagenesis of the 3 key SYR residues using *T.in(exo-)* Pol I again demonstrates the majority of mutations to yield active strand-displacing polymerases. Figure 6.27b highlights the serine residue (blue bar) and tyrosine (red bar) to be active with all but 6 of the possible amino acid mutations. The arginine residue (green bar) again only shows activity with a positive amino acid residue in that position.

The mutagenesis results of the *G.me*\* and *T.in(exo-)* Pol I do not identify any residue to permitting reduced amplification times over the wild-type enzyme: i.e. none of the mutations creates an enzyme with a shorter time to threshold than the WT enzyme.

### 6.3 Discussion

This chapter details the successful engineering of several polymerases by site-directed mutagenesis using overlapping PCR primers. *G.me* Pol I was initially mutated to create a sequence and functional equivalent of *B.st* Pol I (NEB, UK). This enabled all further engineering and characterisation studies to be performed with the commercial gold standard isothermal enzyme. The mutated enzyme *G.me*\* Pol I was shown to be identical to the wild-type *G.me* and *B.st* Pol I in all characterisation studies.

A significant observation from Chapter 3 was the presence of the key amino acid residues potentially required for an active 3'-5' exonuclease domain. *T.in*, *T.at* and *K.ol* Pol I were mutated at Asp<sup>355</sup> and Glu<sup>357</sup>, homologous to the positions in *E.co* Pol I, to alter the amino acid to a small and inert alanine. These mutations had been reported to remove the proof-reading activity in *E.coli* (Derbyshire *et al.*, 1988). The subsequent mutant enzymes, *T.in*(exo-), *T.at*(exo-) and *K.ol*(exo-) Pol I, demonstrated identical 5'-3' DNA polymerase extension rates compared to the wild-type enzymes, confirming the mutations had not affected the ability of the enzyme to incorporate nucleotides onto a pre-annealed ssDNA primer. Mutations did not show a significant effect on the thermal denaturation profile. Confirmation of the reduced ability of the proof-reading domain was observed during the LAMP reaction. *T.in* and *T.at* Pol I showed no significant effect on amplification times with or without the inactivating mutation. Chapter 5 indicated the presence of strand-displacement activity in *K.ol* Pol I in the M13mp18 ssDNA pre-annealed primer extension assay. However, the wild-type enzyme was not found to be suitable for LAMP. It was reasoned the 3'-5' exonuclease activity may have been preventing amplification due to significant degradation of the primers required for amplification. This theory was confirmed. The *K.ol*(exo-) Pol I mutant was shown to amplify in LAMP.

High concentrations of the wild-type enzyme were observed to produce an increasing level of fluorescence prior to a successful LAMP amplification result. This increase in fluorescence was not observed with either the exo- mutants or when the reaction was repeated using phosphorothioated primers, that prevented primer degradation. The ssDNA LAMP primers were therefore reasoned to be degraded by the 3'-5' exonuclease, leading to non-specific annealing to the template DNA, and background linear extension. This observation was more obvious in the presence of *K.ol* Pol I, and required 8-fold less enzyme than either *T.in* or *T.at* Pol I to be observed. This suggested the *T.in* and *T.at* Pol I have significantly weaker 3'-5' exonuclease activity than that of *K.ol* Pol I. *Thermotoga maritima* Pol I has been reported to show strong 3'-5' exonuclease activity (Yang *et al.*, 1999);

therefore, it is not surprising to see the more closely related enzyme to also show this activity. Due to the reduced thermostability and slower amplification time, compared to *T.in(exo-)* Pol I, the *K.ol(exo-)* Pol I enzyme was not carried further in this study.

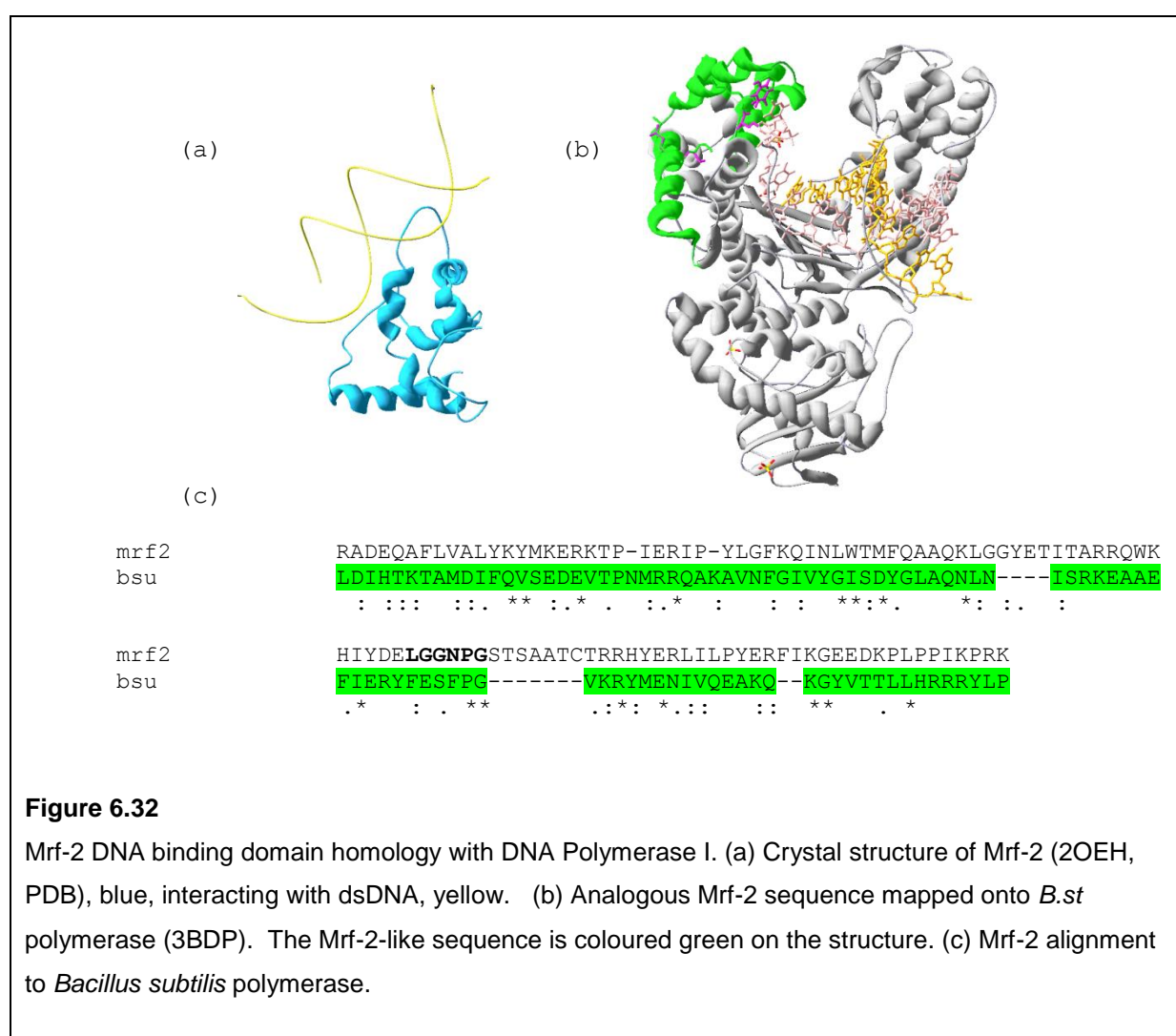
The successful reduction of the 3'-5' exonuclease activities demonstrated the overlap extension site-directed mutagenesis method was an effective method of evolving the polymerases. Further engineering was therefore directed towards those enzymes not showing strand-displacement activity in Chapter 5.

*T.aq* and *T.ma* Pol I are widely used in biotechnology due to their high thermostabilities, making them particularly suitable for PCR reactions. Indeed, *T.aq* Pol I was the first thermostable PCR enzyme to be identified; replacing *E.coli* Klenow polymerase in amplification reactions and enabling the rapid increase in PCR based methods in the 1980s (Saiki *et al.*, 1988).

Family A polymerases are required for the repair of Okazaki fragments during DNA replication. As such, the enzyme must strand displace to enable RNA primer removal. It is suggested that the polymerases demonstrating weak strand-displacement activity *in vitro*, require accessory proteins *in vivo* for efficient activity (Viguear *et al.*, 2001). It was therefore inferred that *B.st* Pol I must have significant activity, not requiring additional proteins for activity *in vivo*. *T.aq* and *T.ma* Pol I were mutated to *B.st*-like residues over the region, reported by Singh *et al.* (2007), to be responsible for strand separation activity. The 5'-3' polymerase activity of the *T.aq*\_SD mutant polymerase was shown to be reduced 4-fold over the wild-type enzyme, and the enzyme did not demonstrate activity in LAMP. The *T.ma* mutant polymerases were both shown to have increased nucleotide incorporation activity over the wild-type enzyme, but again failed to show strand-displacement activity. This mutagenesis result did not elucidate the mechanism of strand-displacement activity as was hoped.

Understanding the reasons for the varying activity observed, directed research towards the limited number of reports investigating strand-displacement activity. Singh *et al.* (2007) have implicated the role of the Fingers subdomain, and specifically a three-helix bundle in the O1-helix, to play a role in strand separation. They demonstrated, using the *E.coli* Klenow Fragment polymerase, that the mutation of three key residues (Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup>) to alanine removed strand-displacement activity. It has also been noted that the structure of the Fingers subdomain, formed by the O, O1 and O2 helices of DNA Pol I, shares structural homology with a novel DNA binding motif (Yuan *et al.*, 1998; Singh *et al.*, 2007; Xie, 2012).

This Mrf-2 DNA binding domain is a member of the AT-rich interaction domain family (ARID) (Yuan *et al.*, 2001) and has a high affinity to specific sequences of dsDNA (Yuan *et al.*, 1998). The Fingers subdomain is therefore inferred to have an affinity to specific dsDNA sequences, driven by the hydrogen bonding interactions between the protein and the specific site of the DNA lattice (Bedinger *et al.*, 1989).



It has been suggested by Xie (2012) that the downstream base pair is unwound by the O1-helix present in the Fingers subdomain as the polymerase moves forward. Sequence-dependent pausing, due to the binding affinity in the Fingers subdomain, is proposed to



increase the probability of cleaving the displaced non-template strand by the 5'-nuclease domain *in vivo*. This result may explain the lack of strand-displacement activity in *T.aq* and *T.ma* polymerases. If the Fingers subdomain does not exhibit the same binding affinity to the ssDNA template compared to a stronger strand-displacing polymerase, they may require accessory proteins to stimulate this activity, as suggested by Viguear *et al.* (2001).

The 3-helical structure, sharing homology to the Mrf-2-like Fingers subdomain, was targeted for mutagenesis to investigate the role of the essential residues, reported by Singh *et al.* (2007). The corresponding Ser<sup>769</sup>, Phe<sup>771</sup> or Arg<sup>841</sup> residues of *B.st* and *T.in(exo-)* Pol I had not been previously investigated for their role in strand-displacement activity.

The three residues were subjected to saturation site-directed mutagenesis, using *G.me\** and *T.in(exo-)* Pol I, enabling all possible residues to be studied at these positions, which had not previously been reported. The use of 'small-intelligent primers' effectively generated the full degeneracy required to obtain all amino acid mutants at each of the three sites for the two enzymes. 120 mutants were identified and over-expressed for characterisation. Singh *et al.* (2007) stated that Ser<sup>769</sup>, Phe<sup>771</sup> and Arg<sup>841</sup> are required for strand-displacement synthesis' in *E.coli* Klenow. This has been shown not to be the case for *G.me* or *T.in(exo-)* Pol I.

The 2 residues (Ser<sup>769</sup> and Phe<sup>771</sup>) located on the 3-helix bundle were shown to be tolerant to a wide variety of mutations. These mutations covered amino acids with varying properties, including positively charged (Arg/R, His/H, Lys/K) and negatively-charged (Asp/D, Glu/E) side-chains. Due to the wide variety of possible substitutions it is difficult to explain the exact role of this residue in terms of its interaction with the downstream template strand. Singh *et al.* (2007) proposed Ser<sup>769</sup> and Phe<sup>771</sup> provide the means to separate the dsDNA structure, with Arg<sup>841</sup> stabilising the template strand.

One clear observation is the requirement for a positive residue at the Arg<sup>841</sup> position, demonstrated to be essential for strand-displacement activity in *G.me\** and *T.in(exo-)* Pol I. It was also noted that *G.me\** Pol I showed faster strand-displacement reactions when a bulky amino acid residue (F, Y, W, R, H, K) was located at the Phe<sup>771</sup> position. Singh *et al.* (2007) demonstrated the Phe<sup>771</sup> is critical in *E.coli* Klenow to act as a physical barrier, separating the dsDNA strands. The results from this study would further support this theory. It is difficult to confer the same conclusions from the *T.in(exo-)* Pol I results, possibly because the reactions take longer, and therefore the differences are more difficult to resolve.

The Fingers subdomain has been implicated and demonstrated to be involved in strand-displacement synthesis but the saturation mutagenesis results in this study suggest further residues and motifs are required in *G.me* and *T.in(exo-)* Pol I, over and above those reported by Singh *et al* (2007).

Chapter 7 will explore these further regions and motifs to localise the exact domains responsible for efficient strand-displacement activity. *T.in(exo-)* Pol I will be used for future study, due to its novel thermostability and strong strand-displacement activities. *G.me*<sup>\*</sup>, the functional equivalent of *B.st* Pol I, will be used as a control enzyme for comparison.

## 7. CHIMERAS

### 7.1 Introduction

The directed evolution by site-directed mutagenesis is one method of protein engineering. An alternative method to evolve polymerase characteristics is to create hybrid enzymes (Nixon *et al.*, 1998), that is, engineered to contain elements of two or more different polymerases.

Structural studies of DNA polymerases have shown that the tertiary arrangement of the Palm, Fingers and Thumb subdomains is highly conserved across all families despite wide variability of amino acid sequences. Using the properties of known enzymes, guided by detailed sequence and structural knowledge, whole domains can be swapped to construct hybrid enzymes.

DNA polymerases have been extensively studied, using T7, *T.aq*, *E.co* and *B.st* DNA polymerases as the classical enzymes with which to compare sequence, structure and function relationships. These data have enabled common residues and motifs to be defined, suggesting shared mechanisms of polymerase activity. The high degree of similarity between Family A polymerases enabled point directed mutations to be engineered, as reported in Chapter 6; with the 3'-5' exonuclease inactivating point mutations providing a high level of confidence that the mutant enzymes will display the desired characteristics.

The SYR single point mutants described in Chapter 6 identified the residues required for strand-displacement activity but also demonstrated the flexibility in terms of the ability to substitute those residues with alternative amino acids. This result, in combination with the strand-displacement characterisation studies reported in Chapter 5, suggest further residues and/or motifs may be required to provide efficient strand-displacement activity to the polymerase.

The creation of hybrid enzymes, to swap polymerase domains, was selected to narrow down and locate regions and domains responsible for the variations in strand-displacement activities observed. A DNA shuffling approach has been reported to create hybrid enzymes (Stemmer *et al.*, 1994). Here, a gene is randomly fragmented and reassembled by error-prone PCR, and the mutants subjected to a round of selection. Further shuffling is repeated until the desired characteristic is successfully produced. A more defined approach was desired for this study, directing the swapping of known regions. It was important to create

seamless joining of the hybrid domains, and an overlap extension method was chosen (Horton, 1989). Splicing by overlap extension PCR enables the creation of chimeras at defined residues. The design and location of the chimeric regions would be critical to minimise unfavourable interactions, and to enable an active enzyme to be created.

DNA polymerase chimeras have been previously investigated. Vilbrandt *et al.* (2000) used domain exchange between *Thermus aquaticus*, *Escherichia coli* and *Thermotoga neapolitana* DNA polymerases. They demonstrated the use of reported crystal structures with which to build a protein scaffold for the design of chimeras to confer increased thermostability and altered processivity to the hybrid enzymes. It was proposed that a similar method could be adopted for this study, using the reported crystal structures for *B.st* and *T.aq* Pol I to identify the domains and boundaries for chimera design.

### 7.1.1 Chapter overview

The creation of chimeras between the novel enzymes identified so far in this thesis may help to identify the exact regions enabling these enhanced activities. Having identified *G.me*<sup>\*</sup> Pol I to amplify rapidly in LAMP, and *T.in(exo-)* Pol I to be highly thermostable, the ultimately desired chimeric protein would have a combination of the two activities. The design of a hybrid enzyme with the amplification speed of *G.me*<sup>\*</sup> Pol I, yet the thermostability of *T.in(exo-)* Pol I would be of significant interest as a commercial enzyme.

## 7.2 Results

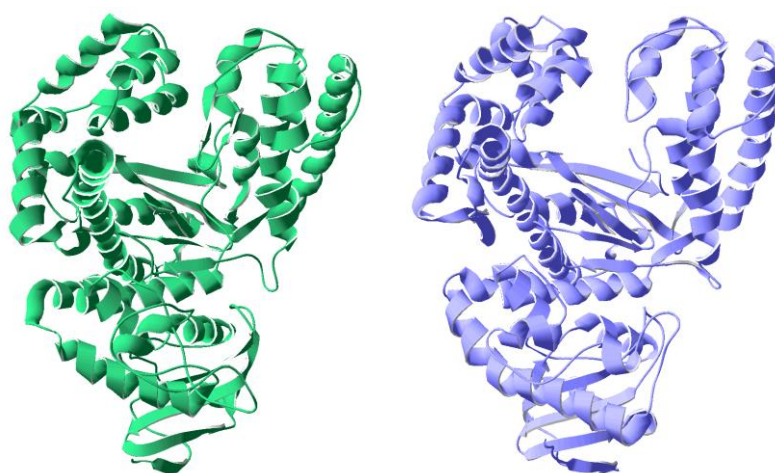
A variety of enzyme chimeras were designed to identify and locate specific activities within different polymerases. The hybrid enzymes include:

- Chimeras between *T.in*(exo-) and *G.me*\* Pol I; with an aim to create a highly thermostable enzyme with the strand-displacement activity of *G.me* Pol I.
- Chimeras between *T.in*(exo-) and *T.aq* Pol I; to create a unique *T.aq*-like enzyme suitable for LAMP.

### 7.2.1 *T.in*(exo-)/*Gme*\* LF DNA polymerase I Chimeras

#### Chimera positioning

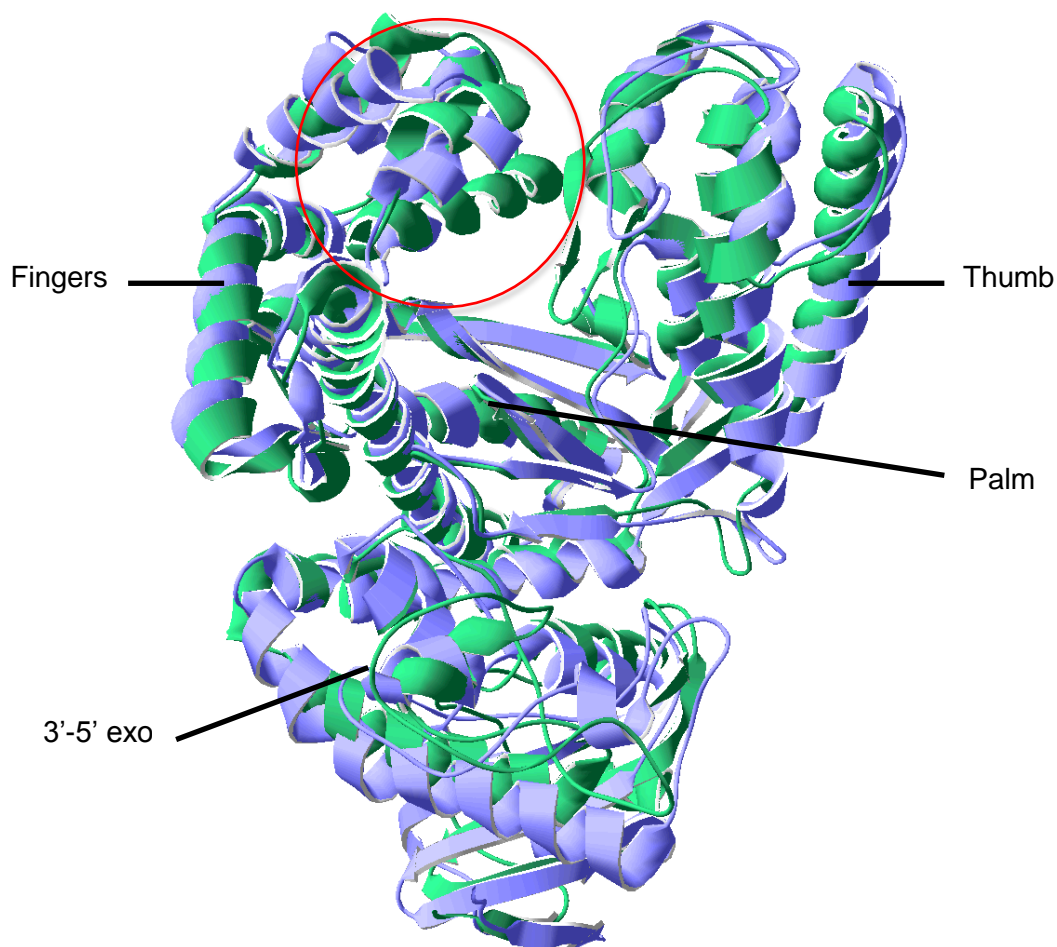
The *T.in*(exo-) Po I amino acid sequence was aligned with the sequence of *KlenTaq* Pol I (4KTQ) (Korolev *et al.*, 1998) using ClustalW. With a sequence identity of 48% it was proposed that 4KTQ could be used as a scaffold to overlay and model the *T.in*(exo-) Pol I sequence. *G.me*\* Pol I was aligned with the sequence of the *B.st* Pol I (3BDP) (Kiefer *et al.*, 1998). With 88% sequence identity a reliable structural model was predicted. The ClustalW alignments were submitted to Swiss-Model enabling homology models to be created using known protein scaffolds (Arnold *et al.*, 2006). The crystal structures identified the standard domains of DNA Pol I-like enzymes, with Thumb, Fingers, Palm and N-terminal 3'-5' exonuclease domains.



**Figure 7.1**

Swiss-Model derived homology models. *T.in*(exo-) Pol I modelled on 4KTQ (green) and *G.me*\* Pol I modelled on 3BDP (blue).

The *T.in(exo-)* and *G.me\** Pol I models were aligned using Swiss-PDBViewer (Guex *et al.*, 2009) to identify significant structural variations. The model identified a shift of the *T.in(exo-)* Pol I O-helix, located in the Fingers subdomain, towards the Palm subdomain. (Figure 7.2). It was proposed that this shift may affect the ability of the two enzymes to strand-displace and was therefore targeted for chimera formation.

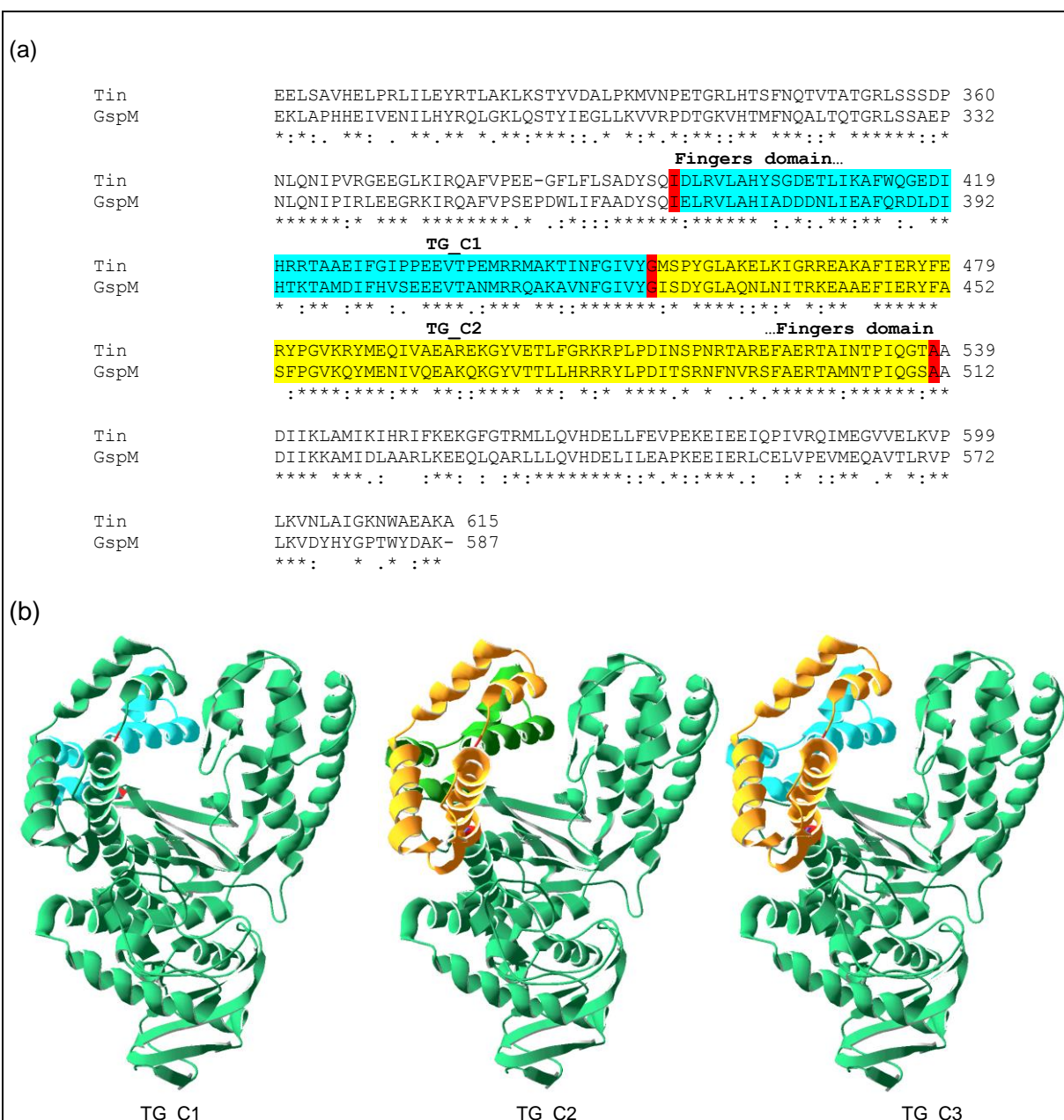


**Figure 7.2**

*T.in(exo-)* Pol I model (green) aligned onto *G.me\** Pol I model (Blue) to compare their tertiary structures. The red circle highlights a significant structural difference, located within the O-helix bundle. The individual subdomains are also located.

The *G.me\** Pol I Fingers subdomain was divided into three regions, based on the crystal structures and sequence alignments with *T.in(exo-)* Pol I. Localised regions with higher sequence identity were chosen for the chimera boundaries to keep potential structural

rearrangements to a minimum (Figure 7.3a). The three chimeras were identified as TG\_C1, TG\_C2 and TG\_C3 (Figure 7.3b). In each case, a small region of the *G.me*<sup>\*</sup> Pol I Finger subdomain sequence was inserted into *T.in(exo-)* Pol I sequence, i.e the predominant sequence remained *T.in(exo-)* *polA*. In this way, it was hoped to keep the thermostability of *T.in(exo-)* Pol I, but introduce the strand-displacement activity of *G.me*<sup>\*</sup> Pol I.



**Figure 7.3**

(a) *T.in(exo-)* Pol I partial alignment with *G.me*<sup>\*</sup> Pol I over the Fingers subdomain region. The conserved chimera boundaries are highlighted red. The three chimeras are denoted: TG\_C1 (light blue), TG\_2 (yellow), TG\_C3 is composed from TG\_C1 and TG\_C2 (Light blue and yellow). (b) Chimera locations highlighted on the *T.in(exo-)* Pol I model (green): TM\_C1 (light blue), TM\_C2 (yellow), TM\_C3 (light blue and yellow). The chimera boundaries are highlighted red.

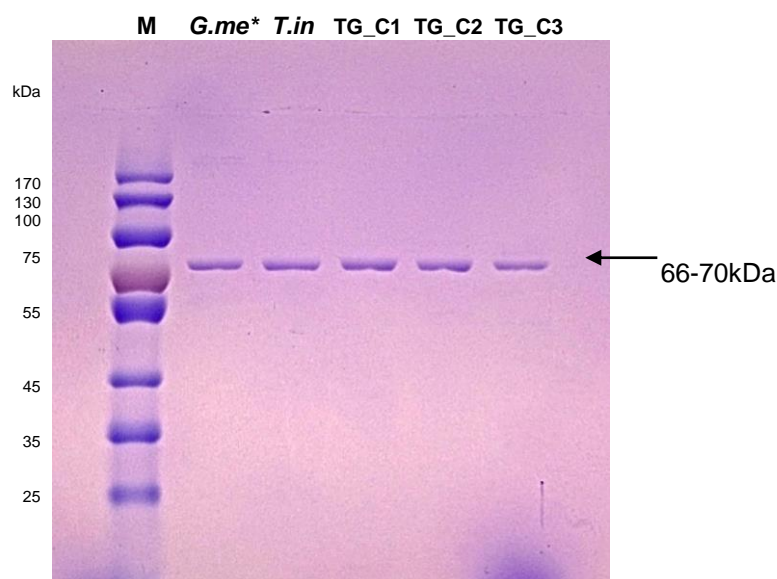
Chimeras were created using the overlap extension PCR method, described in Chapter 2.3.6. using the primer sequences in Appendix iii.i. Table 7.1 details the individual high fidelity PCR reactions required to create each chimera and the plasmid template used. *Dpn* I restriction enzyme treated PCR products from the individual PCR reactions were used as template for the final overlap PCR. The final purified 2075bp PCR fragments were *Nde* I and *Bam*H I digested for directional cloning into pET24a+HIS. Ligations were transformed into *E.coli* KRX(pRARE2). Five colonies of each chimera were PCR screened using T7 primers and DNA sequenced to confirm the sequence of each chimeric *po*/A. A single clone of each chimera was over-expressed on a large-scale and purified as previously reported. 1µg purified enzyme was visualised on a 12% SDS-PAGE gel to confirm the purity (Figure 7.4).

	1 <sup>st</sup> PCR	2 <sup>nd</sup> PCR	3 <sup>rd</sup> PCR	Final PCR
<b>TG_C1</b>	T7promoter TG_1_L1  <i>T.in(exo-)</i> pDNA  1382bp fragment	TG_1_U1 TG_1_L2  <i>G.me*</i> pDNA  217bp fragment	TG_1_U2 T7terminator  <i>T.in(exo-)</i> pDNA  645bp fragment	T7promoter T7terminator  1 <sup>st</sup> + 2 <sup>nd</sup> + 3 <sup>rd</sup> PCR products  2075bp fragment
<b>TG_C2</b>	T7promoter TG_C2_L1  <i>T.in(exo-)</i> pDNA  1474bp fragment	TG_C2_U1 TG_C2_L2  <i>G.me*</i> pDNA  304bp fragment	TG_C2_U2 T7terminator  <i>T.in(exo-)</i> pDNA  392bp fragment	T7promoter T7terminator  1 <sup>st</sup> + 2 <sup>nd</sup> + 3 <sup>rd</sup> PCR products  2075bp fragment
<b>TG_C3</b>	T7promoter TG_C1_L1  <i>T.in(exo-)</i> pDNA  1382bp fragment	TG_C1_U TG_C2_L  <i>G.me*</i> pDNA  472bp fragment	TG_C2_U2 T7terminator  <i>T.in(exo-)</i> pDNA  392bp fragment	T7promoter T7terminator  1 <sup>st</sup> + 2 <sup>nd</sup> + 3 <sup>rd</sup> PCR products  2075bp fragment

**Table 7.1**

TG chimera PCR amplification overview; detailing the primer combinations and pDNA template required to generate each chimeric PCR fragment for cloning. PCR fragment sizes described are inclusive of the vector polylinker region (92bp to T7promoter and 135bp to the T7terminator sequence). Individual chimera primer sequences can be found in Appendix iii.i.





**Figure 7.4**

1ug purified Pol I chimeras were visualised on a 12% SDS-PAGE. Lane 1: M: PageRuler protein ladder, Lane 2: *G.me\** Pol I, Lane 3: *T.in(exo-)* Pol I, Lane 4: TG\_C1 Pol I, Lane 5: TG\_C2 Pol I, Lane 6: TG\_C3 Pol I. The overexpressed Pol I is identified by an arrow.

### Characterisation

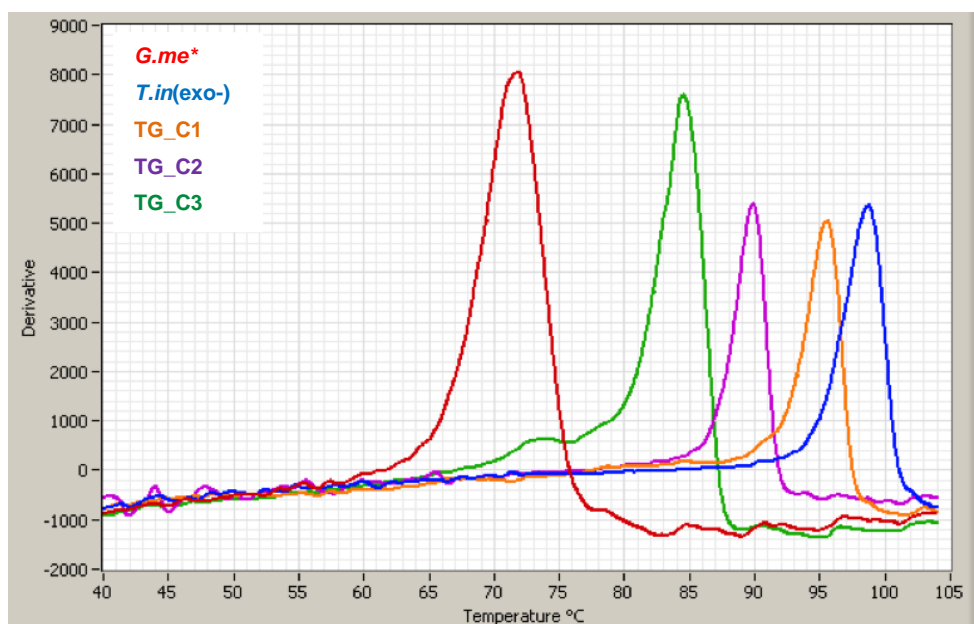
The 5'-3' DNA polymerase extension activity of the three Pol I chimeras were compared to the wild-type *G.me\** and *T.in(exo)* Pol I, as described in Chapter 2.6.10. All chimeras were shown to remain active, with varying extension rates compared to the wild-type enzymes (Table 7.2).

	Specific activity ( <i>B.st</i> -equivalent units/mg)
<i>G.me*</i>	120,000
<i>T.in(exo-)</i>	15,000
TG_C1	27,250
TG_C2	31,250
TG_C3	15,000

**Table 7.2**

Activity confirmation using the M13mp18 ssDNA primer extension assay, using known *B.st* Pol I units as a standard.

With the Pol I chimeras exchanging up to 142 amino acid residues, between *T.in(exo-)* and *G.me\** Pol I (Figure 7.3a), it is expected that the thermostability of the chimeras will have been altered. To test for these variations, 1 µg of each Pol I was denatured in the presence of SYPRO®-Orange on a Genie®II to monitor the thermal denaturation profile of each chimera, as per Chapter 2.6.3. (Figure 7.5 and Table 7.3).



**Figure 7.5**

Thermal Melt profile comparing the denaturation profiles of the chimeras and the wild-type Pol I enzymes. Reactions were reported following the method in Chapter 2.6.3.

	$T_m$ (°C)
<i>G.me*</i>	71.8
<i>T.in(exo-)</i>	98.7
TG_C1	95.7
TG_C2	90.0
TG_C3	84.4

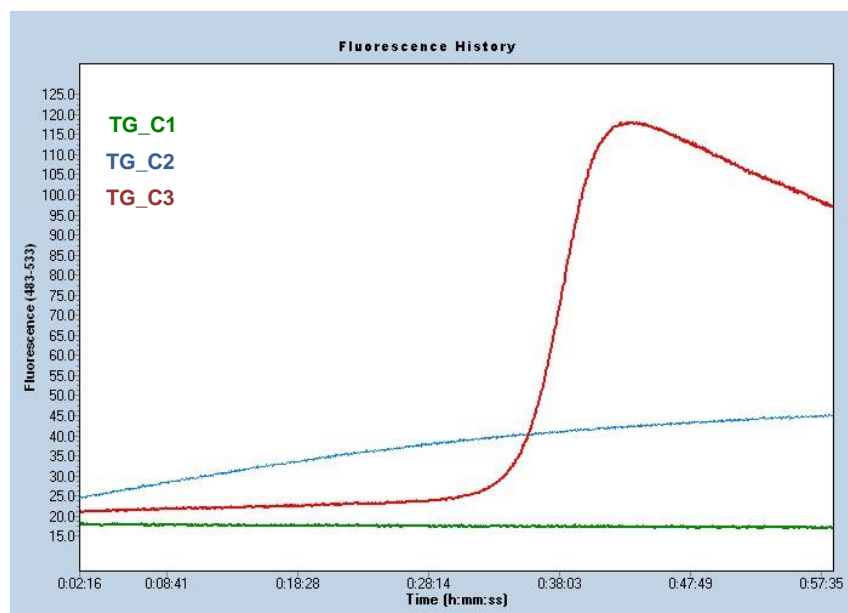
**Table 7.3**

Table comparing the  $T_m$  of each chimera and wild-type polymerase.  $T_m$  is the point of maximal rate of change in fluorescence over temperature.

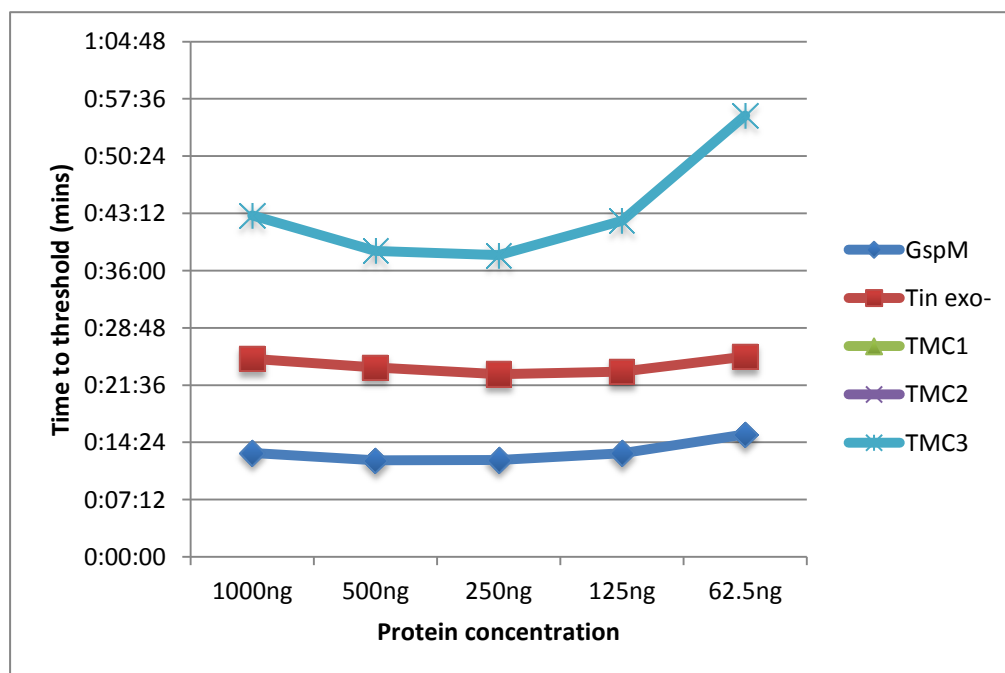
All chimeras were shown to be less thermostable than *T.in(exo-)* Pol I, as shown in Figure 7.5 (Blue plot). TG\_C1 has a 57 amino acid region swap (of which, 24 amino acids are different compared with the *G.me\** Pol I sequence), TG\_C2 has an 84 amino acid region swap (of which, 30 amino acids are different compared with the *G.me\** Pol I sequence) and TG\_C3 has a 142 amino acid region swap (of which, 54 amino acids are different to *G.me\** Pol I). As the proportion of the *G.me\** Pol I sequence increases in each chimera, the thermostability of that chimera decreases. Surprisingly, TG\_C3, with 54 of the 142 amino acid residues in the chimera region of the Fingers subdomain altered, is still thermostable to 84.4°C.

The strand-displacement activity of the chimeras was then investigated and compared to the wild-type enzymes. 10ng *A.pe* gDNA was run in a LAMP reaction, with a dilution of the enzyme chimeras, in comparison to the wild-type enzymes. Reactions were run at 65°C on the LightCycler®-480 as per Chapter 2.6.5. TG\_C3 was shown to be active in LAMP and therefore the chimera region did not remove the ability of the enzyme to strand-displace (Figure 7.6a). The amplification time was shown to increase by 30 minutes with 62.5ng (8u *B.st* Pol I equivalent) protein in the reaction (Figure 7.6b). TG\_C1 and TG\_C2 did not demonstrate activity in LAMP. TG\_C2 was shown to have an increasing background fluorescence, previously described in Chapter 6 using *K.ol(exo-)* enzyme in LAMP. The location of the TG\_C2 chimera was not expected to interact with the 3'-5' exonuclease domain of the *T.in(exo-)* Pol I and therefore the result was not thought to be due to an alteration of the activity of the proof-reading function of the enzyme. Rather, the background fluorescence was thought to be due to enzyme specific interactions with the primers, leading to the formation of dsDNA primer dimers.

(a)



(b)



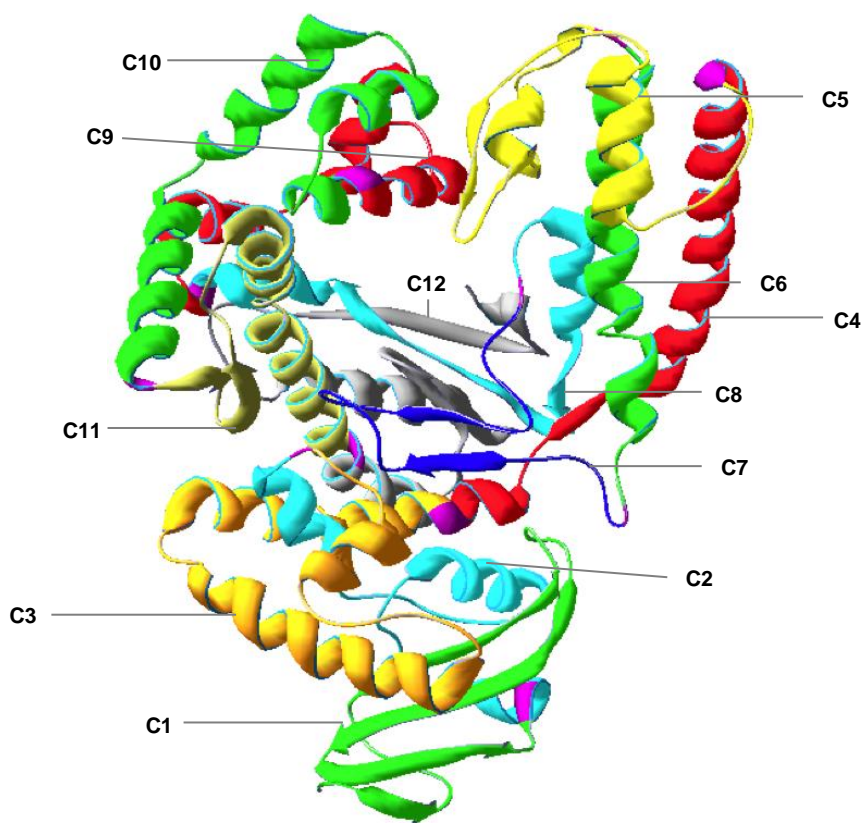
**Figure 7.6**

Activity of the enzyme chimeras in LAMP (a) real-time amplification plot comparing 250ng each chimera: TG\_C1 (green), TG\_C2 (blue), TG\_C3 (red). (b) Amplification times vs protein concentration.

### 7.2.2 *T.in-T.aq* LF DNA polymerase I chimeras

With confidence that a chimeric enzyme can retain activity in LAMP, an additional set of chimeras were designed. The aim of these chimeras was to confer the strand-displacement activity of *T.in*(exo-) Pol I to *T.aq* Pol I, currently lacking the strand-displacing activity required for LAMP. The *T.in*(exo-) Pol I structural model, based on the 4KTQ scaffold in Chapter 7.2.1, was used to define the boundaries of the chimera regions to be made (Figure 7.7). An amino acid alignment confirmed closely related sequences with which to anchor the chimera boundary primers, to keep structural rearrangements to a minimum (Figure 7.8).

Primers were designed to create 18 different chimeras covering the whole length of the LF *T.in*(exo-) and *T.aq* *polA* sequences. 12 chimeras featured a unique region of the *polA* sequence, with 6 further chimeras made up from multiple individual chimeras. Chimera regions were colour coded to highlight their position on the structural model (Figure 7.7).



**Figure 7.7**

*Tin*(exo-) Pol I model structure (based on PDB: 4KTQ); highlighting the location of the 12 individual chimera regions (C1-12), corresponding to the sequences in Figure 7.8.



chimeric PCR fragment. The PCR fragments required for each chimera were *Dpn* I restriction enzyme and were used as template for a final overlap PCR, using the vector specific T7promoter and T7terminator primers. The 18 final high fidelity PCR products, each defining a unique *polA* chimera, were *Nde* I and *Sal* I digested for directional cloning into pET24a+HIS. Ligations were transformed into *E.coli* KRX(pRARE2). Five colonies of each chimera were PCR screened using T7 primers and DNA sequenced to confirm the chimera boundaries were successfully created.

Chimera fragment	pDNA Template	Upper primer	Lower primer
<b>C1_U</b>	Tin(exo-)	T7 promoter	KTaq_Tin(C1)_L
C1_L	Taq	KTaq_Tin(C1)_U	T7 terminator
<b>C2_U</b>	Taq	T7 promoter	KTaq_Tin(C2)_L
C2_I	Tin(exo-)	KTaq_Tin(C2)_U	KTaq_Tin(C2)_L2
C2_L	Taq	KTaq_Tin(C2)_U2	T7 terminator
<b>C3_U</b>	Taq	T7 promoter	KTaq_Tin(C3)_L
C3_I	Tin(exo-)	KTaq_Tin(C3)_U	KTaq_Tin(C3)_L2
C3_L	Taq	KTaq_Tin(C3)_U2	T7 terminator
<b>C4_U</b>	Taq	T7 promoter	KTaq_Tin(C4)_L
C4_I	Tin(exo-)	KTaq_Tin(C4)_U	KTaq_Tin(C4)_L2
C4_L	Taq	KTaq_Tin(C4)_U2	T7 terminator
<b>C5_U</b>	Taq	T7 promoter	KTaq_Tin(C5)_L
C5_I	Tin(exo-)	KTaq_Tin(C5)_U	KTaq_Tin(C5)_L2
C5_L	Taq	KTaq_Tin(C5)_U2	T7 terminator
<b>C6_U</b>	Taq	T7 promoter	KTaq_Tin(C6)_L
C6_I	Tin(exo-)	KTaq_Tin(C6)_U	KTaq_Tin(C6)_L2
C6_L	Taq	KTaq_Tin(C6)_U2	T7 terminator
<b>C7_U</b>	Taq	T7 promoter	KTaq_Tin(C7)_L
C7_I	Tin(exo-)	KTaq_Tin(C7)_U	KTaq_Tin(C7)_L2
C7_L	Taq	KTaq_Tin(C7)_U2	T7 terminator
<b>C8_U</b>	Taq	T7 promoter	KTaq_Tin(C8)_L
C8_I	Tin(exo-)	KTaq_Tin(C8)_U	KTaq_Tin(C8)_U2
C8_L	Taq	KTaq_Tin(C8)_U2	T7 terminator
<b>C9_U</b>	Taq	T7 promoter	KTaq_Tin(C9)_L
C9_I	Tin(exo-)	KTaq_Tin(C9)_U	KTaq_Tin(C9)_L2
C9_L	Taq	KTaq_Tin(C9)_U2	T7 terminator
<b>C10_U</b>	Taq	T7 promoter	KTaq_Tin(C10)_L
C10_I	Tin(exo-)	KTaq_Tin(C10)_U	KTaq_Tin(C10)_L2
C10_L	Taq	KTaq_Tin(C10)_U2	T7 terminator
<b>C11_U</b>	Taq	T7 promoter	KTaq_Tin(C11)_L
C11_I	Tin(exo-)	KTaq_Tin(C11)_U	KTaq_Tin(C11)_L2
C11_L	Taq	KTaq_Tin(C11)_U2	T7 terminator
<b>C12_U</b>	Taq	T7 promoter	KTaq_Tin(C12)_L
C12_L	Tin(exo-)	KTaq_Tin(C12)_U	T7 terminator

**Table 7.4**

A summary of the primers and template combinations required to PCR amplify the relevant chimeric DNA fragments enabling the formation of the 12 chimera regions.

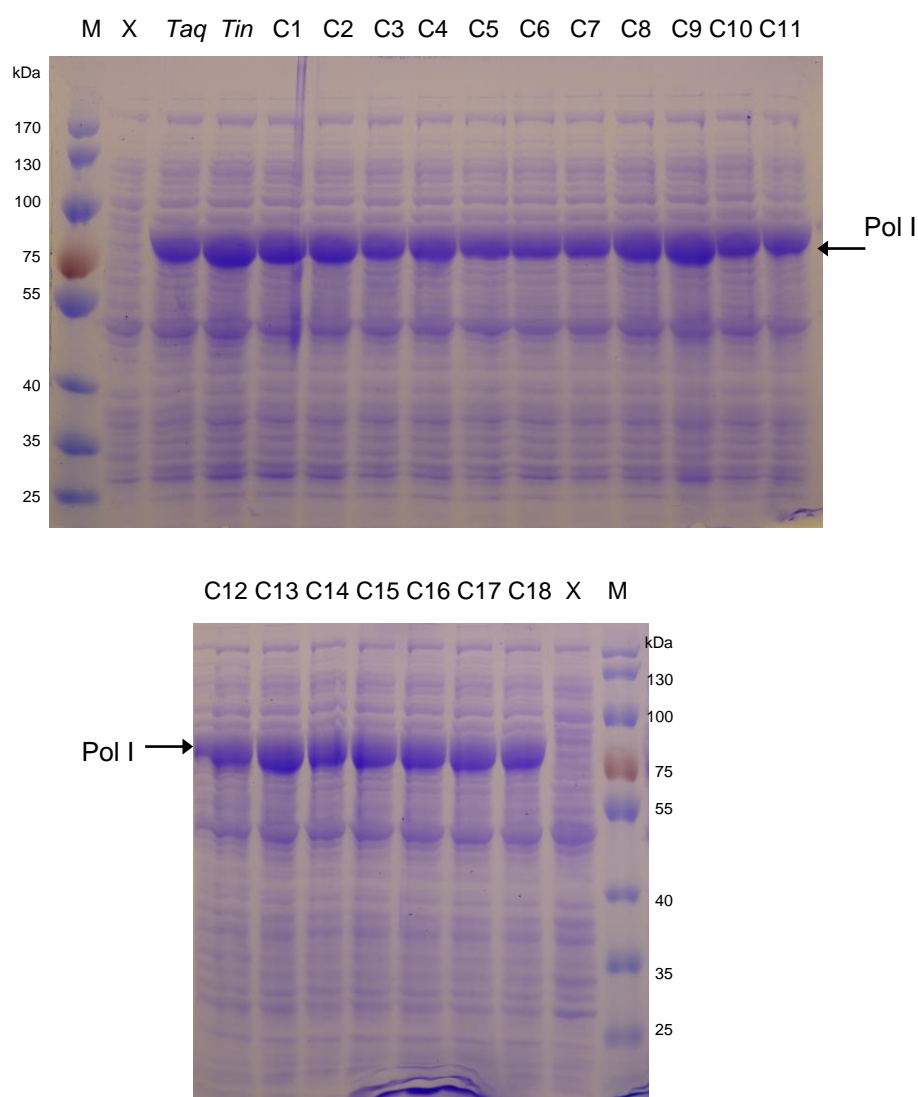
Chimera fragment	Chimeras regions	pDNA Template	Upper primer	Lower primer
<b>C13</b> U	4-6	Taq	T7 promoter	KTaq Tin(C1) L
C13 I		Tin(exo-)		
C13 L	Thumb	Taq	KTaq Tin(C1) U	T7 terminator
<b>C14</b> U	4-12	Taq	T7 promoter	KTaq Tin(C2) L
C14 I	Fingers/	Tin(exo-)	KTaq Tin(C2) U	KTaq Tin(C2) L2
C14 L	Thumb/Palm	Taq	KTaq Tin(C2) U2	T7 terminator
<b>C15</b> U	9+10	Taq	T7 promoter	KTaq Tin(C3) L
C15 I	½ Fingers	Tin(exo-)	KTaq Tin(C3) U	KTaq Tin(C3) L2
C15 L		Taq	KTaq Tin(C3) U2	T7 terminator
<b>C16</b> U	9-11	Taq	T7 promoter	KTaq Tin(C4) L
C16 I	Fingers	Tin(exo-)	KTaq Tin(C4) U	KTaq Tin(C4) L2
C16 L		Taq	KTaq Tin(C4) U2	T7 terminator
<b>C17</b> U	10+11	Taq	T7 promoter	KTaq Tin(C5) L
C5 I	½ Fingers	Tin(exo-)	KTaq Tin(C5) U	KTaq Tin(C5) L2
C5 L		Taq	KTaq Tin(C5) U2	T7 terminator
<b>C18</b> U	4-6 + 9-11	Taq	T7 promoter	KTaq Tin(C6) L
C18 I	Thumb	C13	KTaq Tin(C6) U	KTaq Tin(C6) L2
C18 L		Taq	KTaq Tin(C6) U2	T7 terminator

**Table 7.5**

A summary of the primers and template combinations required to PCR amplify the relevant Chimeric DNA fragments enabling the formation of the 6 additional chimera regions, composed of multiple chimeric regions generated from Table 7.4.



A single clone of each chimera was over-expressed in a 96 deep-well plate format as described in Chapter 2.4.3. Cultures reached an OD<sub>600</sub> of 8.5 after a 22h induction at 37°C. Resuspended samples were heat treated at 65°C for 60min to denature background *E.coli* proteins. The equivalent of 100µl each unfractionated culture was visualised on a 12% SDS-PAGE gel to confirm expression and compare protein yields (Figure 7.9).



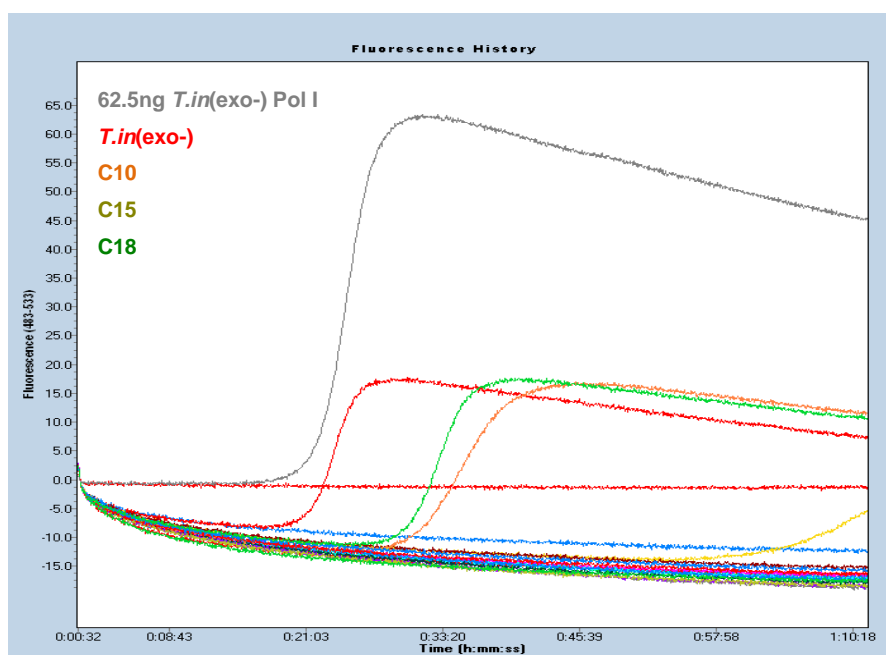
**Figure 7.9**

12% SDS PAGE gel comparing expression yields of the *Tin-Taq* chimeras. Each lane is the equivalent of 100µl culture. M: P protein standard, X: uninduced control, *Taq*: wild-type *Taq* Pol I, *Tin*: wild-type *T.in(exo-)* Pol I, C1-C18: chimeras 1-18. The overexpressed Pol I is highlighted by an arrow.

## Characterisation

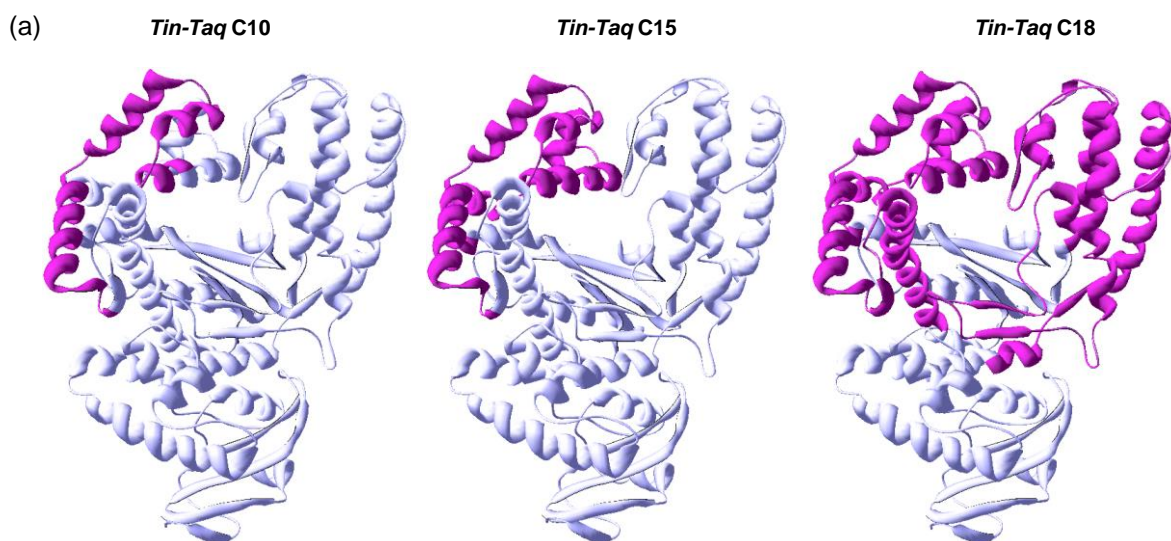
The protein gels confirmed all Pol I chimeras were soluble and expressed to give a similar yield of protein when compared to the wild-type enzyme. A dilution of the unfractionated sonicated sample (2 $\mu$ l culture equivalent) was added directly to a variety of reactions to compare activities to the wild-type enzyme.

Chimeras were run in a 65°C LAMP reaction to investigate strand-displacement activity, as per Chapter 2.6.5. Three chimeras were shown to be positive for strand-displacement activity; that is, C10, C15 and C18 were shown to amplify 10ng *A.pe* gDNA under 1hr. The unfractionated nature of the sample shows a reduced maximum fluorescence value for the samples when compared to the amplification using a purified enzyme stock (Figure 7.10, grey plot). Un purified background *E.coli* nucleic-acids within the extract were expected to be interacting with the intercalating fluorescent dye reducing the dye available to the dsDNA generated during LAMP amplification. This was not expected to alter the activity of the polymerase in the reaction.



**Figure 7.10**

A LAMP reaction performed at 65°C run and on an LC-480 to compare the *Tin-Taq* chimeras. 2 $\mu$ l of a 1 in 10 dilution of sonicated cell extract was used directly in a 25 $\mu$ l LAMP reaction. *T.in(exo-)* Pol I (red), C18 Pol I (green), C10 Pol I (orange), C15 Pol I (yellow). The remaining chimeras and uninduced control are shown by the remaining colours and are seen as flat lined results. 62.5ng *T.in(exo-)* Pol I was used as a control (grey).



(c)

***Tin-Taq C10***

Mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrlldvaylralslevae  
iarleaeavfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlhhprtgrl  
htrfnqtatatgrlsssdnlnipvrtpglgrrirrafiieegwllvaldysqielrvla[ysgdetlikafwggedihrrtaaeifgippeev  
tpemrrmaktinfgivygmspyglakelkigrreakafieryferypgvkrymeqivaearekgyvetlfg[rryvpdlearvksvreaaerma  
fnmpvggtaadlmklamvklfprleemgarmllqvhdelvleapkeriaeavarlakevmegvyplavplevevgigedwlsake

***Tin-Taq C15***

Mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavl[hmvmgikidataylrslremaek  
lkeleekiytlagekfnnsskqlgqilfeklklptvkktpkktaystdnevleelsavhelprlileyrtlaklkstyvdalpkmvn[rtgrl  
htrfnqtatatgrlsssdnlnipvrtpglgrrirrafiieegwllvaldysqielrvla[ysgdetlikafwggedihrrtaaeifgippeev  
tpemrrmaktinfgivygmspyglakelkigrreakafieryferypgvkrymeqivaearekgyvetlfg[rryvpdlearvksvreaaerma  
fnmpvggtaadlmklamvklfprleemgarmllqvhdelvleapkeriaeavarlakevmegvyplavplevevgigedwlsake

***Tin-Taq C18***

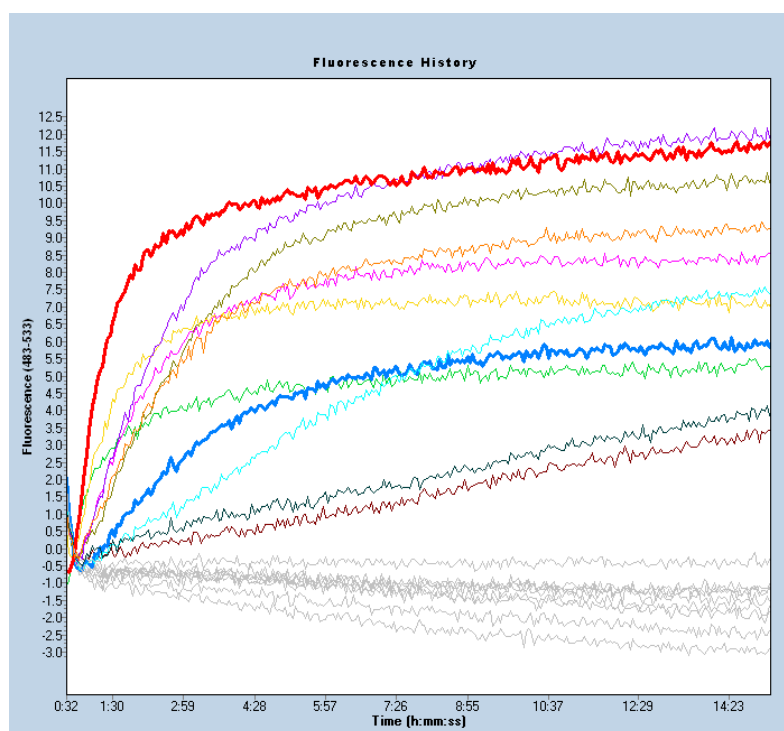
mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavl[hmvmgikidataylrslremaek  
lkeleekiytlagekfnnsskqlgqilfeklklptvkktpkktaystdnevleelsavhelprlileyrtlaklkstyvdalpkmvn[rtgrl  
htrfnqtatatgrlsssdnlnipvrtpglgrrirrafiieegwllvaldysqielrvla[ysgdetlikafwggedihrrtaaeifgippeev  
tpemrrmaktinfgivygmspyglakelkigrreakafieryferypgvkrymeqivaearekgyvetlfg[rryvpdlearvksvreaaerma  
intpiggtadiikla[ivklfprleemgarmllqvhdelvleapkeriaeavarlakevmegvyplavplevevgigedwlsake

**Figure 7.11**

Strand-displacement active chimeras: *Tin-Taq C10,15,18*. (a) Chimeric Pol I model structures. *T.aa* Pol I (pale blue structure), *T.in(exo)* Pol I (pink structure). (b) Chimera locations within the amino acid sequence. Chimeric *T.in(exo-)* DNA Pol I sequence (grey). *T.aa* Pol I sequence (unhighlighted). Chimera boundaries (purple). The remaining chimeras can be seen in the Appendix iii.ii.

Figure 7.11 identifies the locations of the strand-displacement active chimeras. The chimera sequences were homology modelled with 4KTQ using Swiss-model as before. The chimera regions (*T.in(exo-)* Pol I sequence) are highlighted in pink for each chimera to identify their locations within the models for comparison. All chimeras can be seen to share a region of the Fingers subdomain in common, suggesting this region is responsible for conferring strand-displacement activity to *T.aq* Pol I.

With the Fingers subdomain essential for correct nucleotide incorporation in the nucleotidyl transfer reaction (Waksman *et al.*, 2005), it was necessary to confirm the chimeras had not directly affected the 5'-3' DNA polymerase activity of the polymerase. A 1 in 10 dilution of 2µl of the unfractionated cell extract was run in an M13mp18 ssDNA primer extension assay (Figure 7.12). The precise protein concentration within the samples was not known and so it would not be possible to directly quantify the activity. An increase in fluorescence above that of the un-induced control sample indicated the presence of an active 5'-3' DNA Polymerase.



**Figure 7.12**

M13mp18 ssDNA primer extension assay. 2µl of unfractionated cell extracts were run in 50µl reaction at 65°C following Chapter 2.6.10. Wild-type *T.aq* Pol I (blue bold), *T.in(exo-)* Pol I (red bold), un-induced control (light grey), C3 (green), C4 (pink), C6 (yellow), C7 (light blue), C8 (dark green), C10 (gold), C13 (brown), C16 (orange), C18 (purple). The remaining, inactive, chimeras are in light grey.

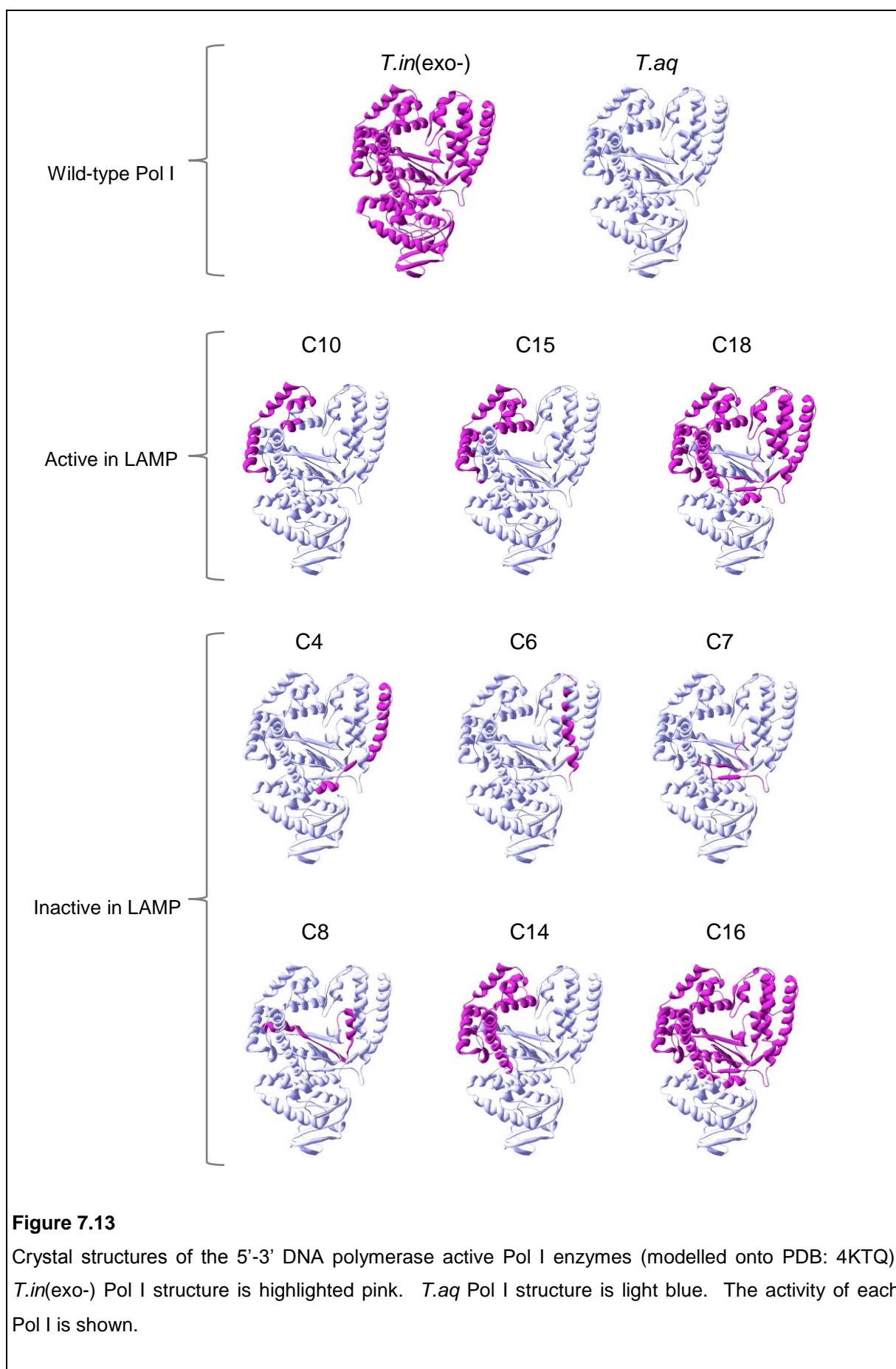
Nine chimeras were shown to retain 5'-3' DNA polymerase activity using the primer extension assay. Chimeras C10, C15, and C18 all show high rates of incorporation (purple, gold and orange plots in Figure 7.12). These chimeras also demonstrate strand-displacement activity in LAMP (Figure 7.10). Chimeras C4, C6, C7, C8, C14, and C16 showed an increase in fluorescence with primer extension, confirming they are active. However, strand-displacement activity in these mutants was not observed (Figure 7.10). Table 7.6 summarises the activity of each 5'-3' polymerase active *Tin-Taq* chimera for comparison; with the location of each, within the *Taq* Pol I crystal structure, detailed in Figure 7.13.

Results suggest the Fingers subdomain is critical for conferring strand-displacement activity from *T.in(exo-)* to *Taq* Pol I. Interestingly, chimeras C10 and C15 individually strand-displace, yet when combined to create C16, this activity is lost.

Pol I	Strand displacement activity	5'-3' DNA polymerase activity
<i>Taq</i>		+
<i>T.in(exo-)</i>	+	+
C1		
C2		
C3		
C4		+
C5		
C6		+
C7		+
C8		+
C9		
C10	+	+
C11		
C12		
C13		
C14		+
C15	+	+
C16		+
C17		
C18	+	+

**Table 7.6**

A comparison of the activities of each 5'-3' polymerase active *Tin-Taq* chimera.



### 7.3 Discussion

Chapter 6 reported the Fingers subdomain to be involved in the strand-displacement activity observed in polymerases from Family A. Whether additional domains were required for this activity was yet to be confirmed. In a search to localise additional residues and domains responsible for the increased strand-displacement activity of *G.me*\* Pol I, a variety of chimeras between the LF DNA Pol I enzymes were designed. With Pol I sequence alignments showing a high degree of similarity between distantly related Pol I enzymes, the shuffling of domains and formation of chimeras was anticipated to offer a unique approach to study strand-displacement activity.

The design of the hybrid enzymes required careful positioning of the chimera boundaries to ensure unfavourable structural rearrangements were kept to a minimum. Villbrandt *et al.* (1997) previously reported the successful formation of chimeras between *T.aq* and *T.ne* Pol I, confirming that thermostable enzyme chimeras can result in active enzymes suggesting chimera formation was therefore a suitable method for this study.

The results in Chapter 5 raised the question of why strand-displacement activity varies between different Pol I enzymes. The *Geobacillus* polymerases were shown to be the most efficient in LAMP reactions, demonstrating strong strand-displacement with a highly active 5'-3' DNA polymerase domain. *T.in(exo-)* Pol I was also shown to be suitable for LAMP, with good strand-displacement activity. A highly thermostable, strand-displacing bacterial polymerase, suitable for isothermal amplification, has not previously been reported in the literature.

The activity of *T.in(exo-)* Pol I was further investigated through the formation of hybrid enzymes with *G.me*\* Pol I, the sequence and functional equivalent to *B.st* Pol I. It was hoped the chimeras would further enhance the amplification speed of the *T.in* enzyme, whilst retaining its high thermostability.

The polymerase Fingers subdomain was initially targeted because it contained two of the three key strand-displacement residues identified by Singh *et al.* (2007). The third key residue (Arg<sup>841</sup>) also interacts directly with the Fingers subdomain. Structural homology modelling, using the reported crystal structures for *T.aq* and *B.st* Pol I (Li, 1998; Kiefer, 1998) as a protein scaffold, provided sites to define the boundaries of the proposed chimeras; these were assigned over regions of high sequence homology. The Fingers subdomain of *G.me*\* Pol I was divided into three regions: (i) the front (DNA interacting), (ii)



the back, and (iii) the front and back. The corresponding regions of *G.me\** *polA* sequences were cloned into the homologous positions of *T.in(exo-)* Pol I to create three different chimeras (TG\_C1, TG\_C2, TG\_C3). The enzymes were shown to be fully soluble, and were expressed in similar yields as the wild-type *T.in(exo-)* Pol I. All chimeras were shown to be active using the 5'-3' DNA polymerase primer extension assay, with two showing enhanced activity over the wild-type enzyme. The chimeras therefore did not significantly alter the homologous catalytic binding pocket in the Palm or the action of the essential Tyr<sup>671</sup> (*E.co* Pol I) permitting nucleotide insertion (Waksman *et al.*, 2005).

The chimeras were all shown to have a reduced thermostability. The denaturation temperature reduced as the proportion of *G.me\** Pol I sequence increased. Less thermostable proteins have not evolved the same thermo-resistant properties and it was reasoned that significant stabilising interactions were lost within the chimeras.

Chimera TG\_C3, a combination of chimeras TG\_C1 and TG\_C2, was shown to be active in LAMP. This result suggests that interactions within the Fingers subdomain between the chimeras are essential for the correct spatial localisation of the Fingers subdomain as a whole. The active chimera had reduced LAMP activity compared to the wild-type enzyme, but retained the same 5'-3' DNA extension activity. This result suggests *T.in* Pol I (15,000 u/mg) strand-displaces more efficiently than the *G.me\** Pol I (120,000 u/mg). The LAMP activity of *T.in(exo)* Pol I is therefore masked by its weaker nucleotide incorporation efficiency. This result further confirmed the results of Singh *et al.* (2007), implying the Fingers subdomain to be crucial for strand-displacement activity.

With the confidence that a chimera between two distantly related enzymes could remain functionally active, additional attempts were made to localise the regions responsible for strand-displacement activity. Alternative polymerase domains, and their use in LAMP, were further explored.

Chapter 5 reported that *T.aq* Pol I does not show the strand-displacement activity required for LAMP. Results in chapter 6 failed to provide a strand-displacing *T.aq* Pol I mutant but the result importantly showed that the presence of the three key strand-displacement residues were not a pre-determinant sequence feature for a strand-displacing polymerase, as reported (Singh *et al.*, 2007). Further residues and motifs must therefore be required for activity and their localisation was investigated through the formation of further chimeras.

Chimeras were designed between *T.aq* and *T.in(exo-)* Pol I. The similar thermostability



profile of the enzymes, demonstrated in Chapter 5, implied chimeras between the two enzymes would be likely to result in active enzymes. 12 regions of the *T.aq* Pol I were identified through structural homology modelling. The regions were individually exchanged with *T.in(exo-)* Pol I sequence with an aim to confer the strand-displacement activity of *T.in* to *Taq*. These chimeras were then used to make additional chimeras, enabling multiple regions to be represented on the *T.aq* Pol I. 18 *Tin-Taq* Pol I chimeras were created, covering the whole length of the LF *T.aq* Pol I sequence.

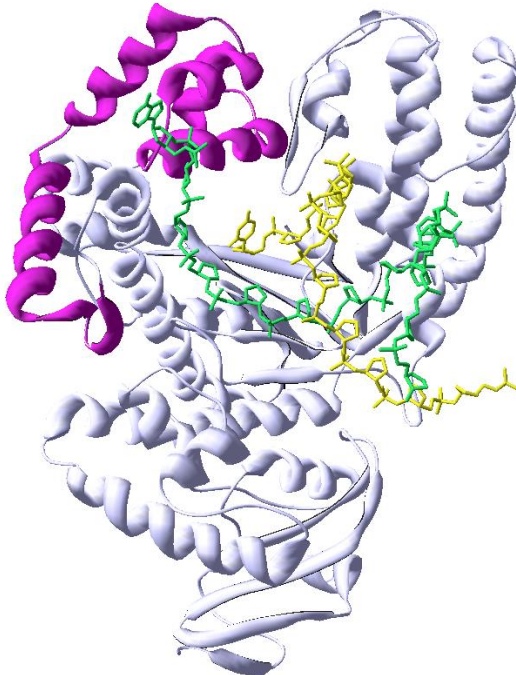
Several chimeras resulted in the loss of 5'-3' DNA polymerase activity, presumably through unfavourable structural rearrangements effecting the catalytic activity of nucleotide incorporation. Of the 9 active chimeras, 3 were seen to be active in LAMP. Structural modelling identified each of these hybrids to contain Fingers subdomain regions of the *T.in(exo-)* Pol I sequence. Interestingly, not all chimeras in the Fingers domain showed LAMP activity. Chimeras 14 and 16 showed 5'-3' DNA polymerase activity but were not active in LAMP. This result mirrors the earlier chimeras between *T.in(exo-)* and *G.me\** Pol I in the Fingers subdomain. This implies unfavourable protein interactions are occurring to prevent the correct dsDNA contacts required for strand-displacement activity. No activity in LAMP was observed by chimeras covering exclusively the Thumb subdomain or Palm subdomain, whole or part, or from the N or C-terminal domains. These domains are therefore not involved in strand-displacement activity.

Chimera 18, covering the Fingers, Thumb and Palm subdomains responsible for the major catalytic and primer/template binding events, was reported to strand-displace. This chimera was also shown to be the most active in the primer extension assay, significantly above that of wild-type *T.aq* Pol I. This result suggests sufficient structural and residue requirements were present from the *T.in(exo-)* Pol I enzyme to confer its higher extension activity to this chimera.

This chapter clearly shows the creation of chimeras between DNA polymerases are possible. Activities have been retained, enhanced or impaired depending on the structural interactions within each chimera. The significant result from this chapter is the exclusive requirement of the Fingers subdomain for strand-displacement activity of DNA polymerase I. The Fingers subdomain therefore required further examination.

Attention was drawn to the significance of the homologous Mrf-2 DNA binding domain (Yuan *et al.*, 1998), mentioned in Chapter 6.3. It was proposed that a Fingers subdomain with increased homology to the Mrf-2 domain would show increased strand-displacement activity.

(b)



(a)

**Mrf-2** RADEQAFLLVALYKYMKERKTP-IERIP-YLGFQKQINLWTFQAAQKLG-GYETITAR-RQWK  
Gme\* LDIHTKTAMDIFHVSEEEVTANMRRQAKAVNFQIVYGISDYGLAQNLN-----ITRK-EAAE  
Tin EDIHRRTAAEIFGIPPEEVTPEMRRMAKTINFGIVYGMSPYGLAKELK--IGRREAK-AFIE  
Taq RDIHTETASWMFGVPRREAVDPLMRRAAKTINFGVLYGMSAHRLSQELAIPYEEAQAIFIERYF

. : :: : \*. \* . : \* . : : : : : \*\* : . :  
. : . : . \* . \* : \* . : : : : : \* : \* :  
\* . : : : \* \* : \* . : : : : : \* : \* : \*

**Mrf-2** HIYDELGCNPGSTSAATCTRRHYERLILPYERFIKGEEDKPLPPIKPRK  
Gme\* FIERYFASFPG-----VKQYMEINIVQEAKQ--KGYVTTLHRRRYLP  
Tin RYFERYPGVKRYMEQIVAEAREKG-----YVETLFGRK-RPLP-----  
Taq QSFVKVRAWIEKT-LEEGRRRGYVETLFGRRRYVP-----

. \* : . . \*\* : : : \* : : : \* : :  
: : . \* . : . \* : : \* : : : : :  
: : : . \* : : \* : : : : : : : \*

**Figure 7.14**

(a) Mrf-2 homology to the Fingers subdomain of the strand-displacing *G.me*\* and *T.in* pPol I, and the non-displacing *T.aq* Pol I. Highly conserved residues are highlighted grey, and the key Ser and Phe/Tyr residues (Singh *et al.*, 2007) are highlighted yellow. (b) The corresponding sequence was mapped onto the *T.aq* Pol I crystal structure 4KTQ for localisation (purple).

Polymerases were shown to align to the Mrf-2 DNA binding domain, confirming the observation by Yuan *et al.* (1998). The domain can be observed to span the key residues reported by Singh *et al.* (2007). *G.me*\*, *T.in* and *T.aq* Pol I appear to share a similar identity to the Mrf-2 sequence. The alignment identifies residues that vary between the polymerase. It is reasoned these sequence variations may lead to significant structural differences, which in turn may affect the efficiency of strand-displacement. This result, in light of the reports from Yuan *et al.*, (1998), Singh *et al.* (2007), and Xie (2012), confirms that the Mrf-2-like DNA binding region, within the Fingers subdomain, is important for strand-displacement activity and will serve as an interesting target for additional study.

This chapter has successfully demonstrated the use of chimeras as a means to localise specific activities of the polymerase, thereby increasing understanding of their mechanisms. *T.in(exo-)* Pol I has so far proved to be the most exciting alternative to *B.st* DNA Pol I for isothermal DNA amplification reactions, but this chapter has failed to provide further enhancements, in terms of faster amplification times. An obvious alternative to engineering, and direct modification, was to employ accessory proteins to enhance the amplification activity of the polymerase. The use of accessory proteins will be investigated in the Chapter 8.

## 8. FUSIONS

### 8.1 Introduction

Polymerase modification and enhancements are continually desired due to the ever increasing needs of the biotechnology sector. Chapter 6 identified the use of site-directed mutagenesis, and Chapter 7, the use of hybrid enzymes (chimeras) to artificially evolve proteins, widening their potential uses. A further enhancement to DNA polymerases is the fusion of an accessory protein to increase the processivity of the enzyme, and this is explored in the current chapter.

The processivity of a DNA polymerase is defined as the average number of nucleotides added to a primer by a single polymerase per association/dissociation step. Loeb *et al.* (2003) demonstrated the fusion of a thioredoxin-binding domain of T3 bacteriophage DNA Pol I at a similar position in the Thumb domain of *T.aq* Pol I. The thioredoxin domain was shown to increase processivity, acting as a clamp. Wang *et al.* (2004) successfully fused a *Sulfolobus solfataricus* (*S.so*) double-stranded DNA binding domain Sso7d to *T.aq* and *Pyrococcus furiosus* (*P.fu*) DNA polymerases. The binding protein is non-sequence specific and was shown to enhance the processivity of the polymerases without affecting the catalytic activity of the enzyme.

Further advantages of fusing DNA binding domains to a polymerase is that they have been reported to increase protein thermostability, and to retain high processivity at increased concentrations of salts and other inhibitors of DNA synthesis, including blood and DNA intercalating dyes (Pavlov *et al.*, 2002). Salt tolerance is favourable when the quality of a template is variable because of salt carry-over from sample preparation (Hamilton *et al.*, 2001). Pavlov *et al.* (2012) further demonstrated the fusion of multiple helix-hairpin-helix (HhH) DNA binding domains of topoisomerase V (Topo V) of *Methanopyrus kandleri* to *T.aq*, *P.fu* and the Large Fragment of *B.st* DNA polymerase. The fusion of the HhH domain to the C-terminus of *B.st* DNA Pol I was shown to increase the half-life of the protein fusion at high temperature, 8-fold compared to that of the wild-type enzyme.

Huang *et al.* (2008) identified a novel family of chromatin proteins that are highly conserved in the thermophilic and hyperthermophilic Crenarchaeota. Cren7 is a small (7kDa) monomeric protein that binds to DNA with no sequence specificity. These DNA binding proteins were shown to significantly improve the processivity of DNA polymerases for use in PCR (Clark *et al.*, 2009).

DNA polymerases containing a fusion protein have been reported for isothermal DNA polymerases. Salas *et al.* (2010) enhanced the Bacteriophage Phi29 DNA polymerase through binding an HhH domain to its C-terminus. This was reported to increase the binding affinity of the hybrid polymerase without affecting the replication rate. Phi29 DNA polymerase is a highly processive enzyme, with significant strand-displacement activity, and is widely used for whole genome amplification using random hexamer/nonamer primers. However, this enzyme shows optimal activity at 30°C and is therefore not suitable for the higher temperature, primer specific isothermal reactions required for nucleic acid detection applications, including LAMP.

The addition of accessory proteins and their advantageous effects have also been demonstrated in LAMP. Single-stranded binding proteins (SSBs) have been reported to increase primer specificity in LAMP reactions through non-sequence specific interactions with the primers at room temperature. These interactions prevent undesired DNA polymerase dependent extension at temperatures lower than the amplification temperature (Tanner *et al.*, 2013).

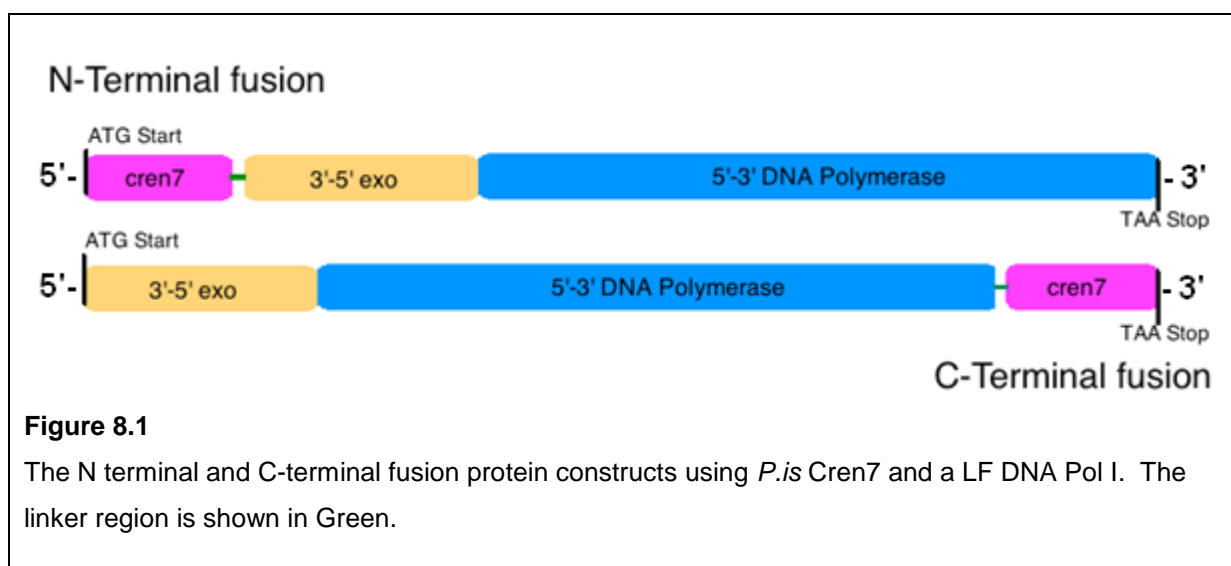
### **8.1.1 Chapter overview**

The use of a DNA binding protein, fused to a DNA polymerase for use in LAMP, has not yet been reported in the literature. As demonstrated in Chapter 6, the overlap extension PCR method can be used to create hybrid enzymes. The method can be further applied to create end-to-end gene fusions (Nixon *et al.*, 1998) enabling the dsDNA binding protein and the polymerase to be expressed in a single open reading frame.

The effects of dsDNA binding proteins, fused to either *G.me*\* or *T.in(exo-)* Pol I will be explored in this chapter, with the aim to create a fusion protein with enhanced characteristics over the wild-type enzyme in LAMP.

## 8.2 Results

Overlapping primers were designed to fuse a Cren7 DNA binding protein to the Large Fragment *G.me*\* and *T.in*(exo-) DNA Pol I. N- and C-terminal fusions were created to compare the effects on each DNA polymerase. A clone containing the highly thermostable archaeal *Pyrobaculum islandicum* (*P.is*) Cren7, optimised for expression in *E.coli*, was kindly donated by Dr Duncan Clark (GeneSys Biotech Ltd.) for this work. Fusion proteins were designed to incorporate the *P.is* Cren7 protein and the LF DNA Pol I, in a single open reading frame, to enable the expression of a single construct containing two active enzymes; separated by a small GTH linker region as reported by Wang *et al.* (2004). The desired fusion constructs are illustrated in Figure 8.1.

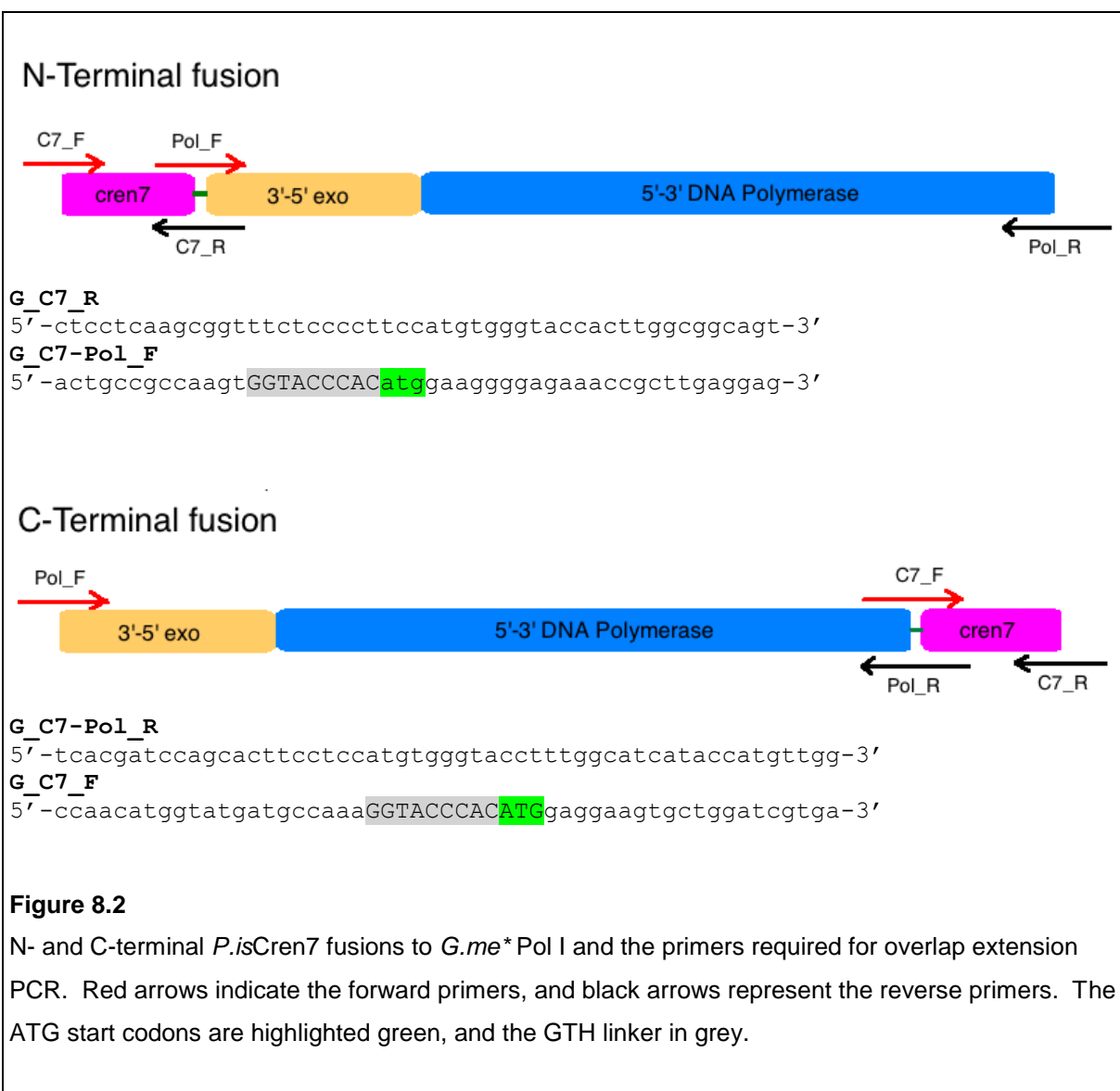


### 8.2.1 Cloning, expression and purification

Two high fidelity overlap PCR reactions were used to generate each of the N- and C-terminal fusions, following the method described in Chapter 2.3.6. The pET24a+HIS\_*Pis*Cren7 clone was available in reading frame from the ATG start codon, as used for all Pol I enzymes used in this study, using the *Nde* I site. The clone could therefore be amplified using the vector specific T7 primers to ensure the *P.is* Cren7 remained in the correct reading frame at the N-terminus.

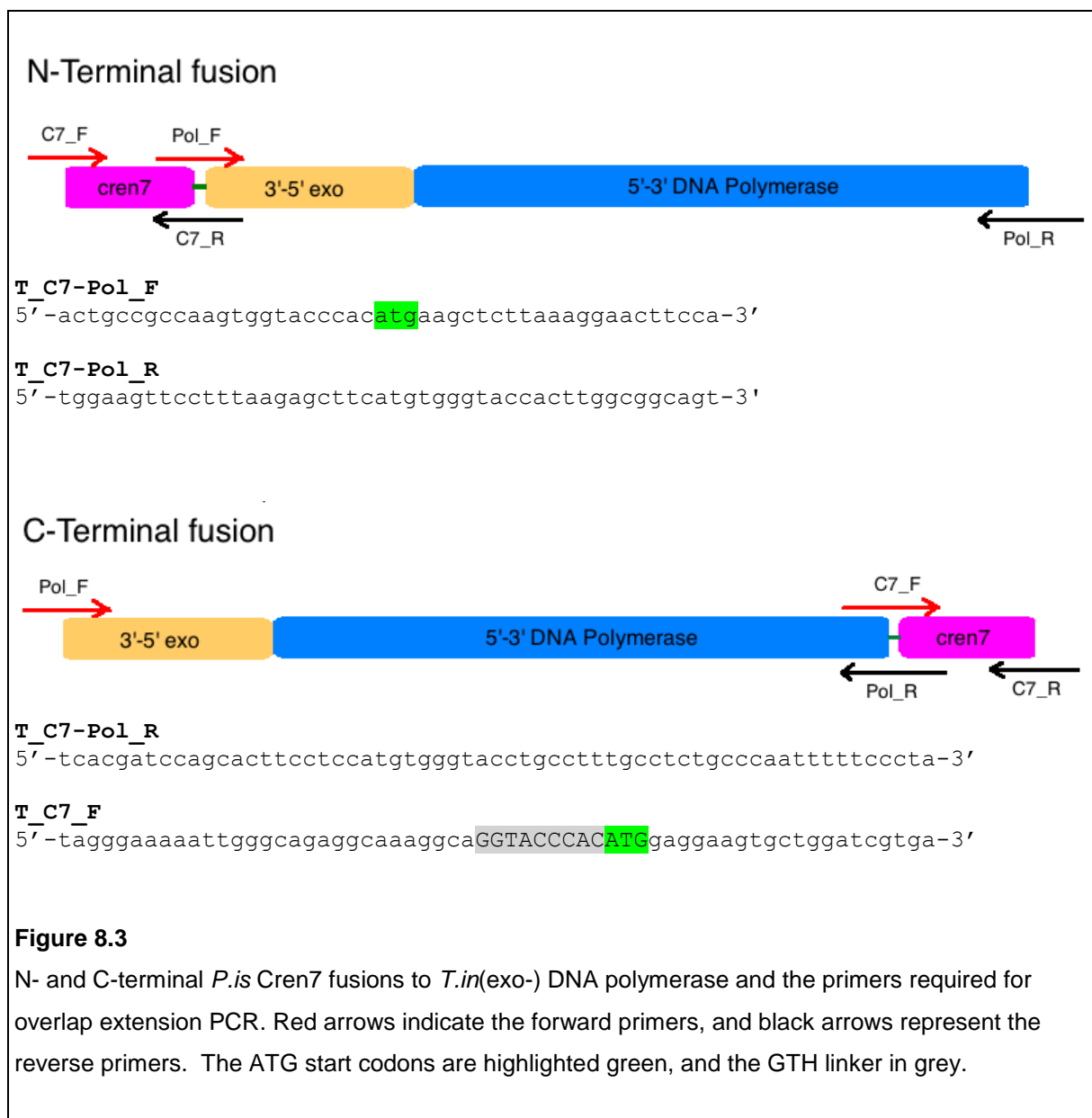
### *P.is* Cren7-*G.me*\* LF DNA polymerase I

The overlap extension PCR process was used, as previously reported in Chapter 6. The specific primers required to generate the fusion proteins are detailed in Appendix iv. An overview of the overlap extension PCR, and the location of the primers required, are detailed in Figure 8.2.



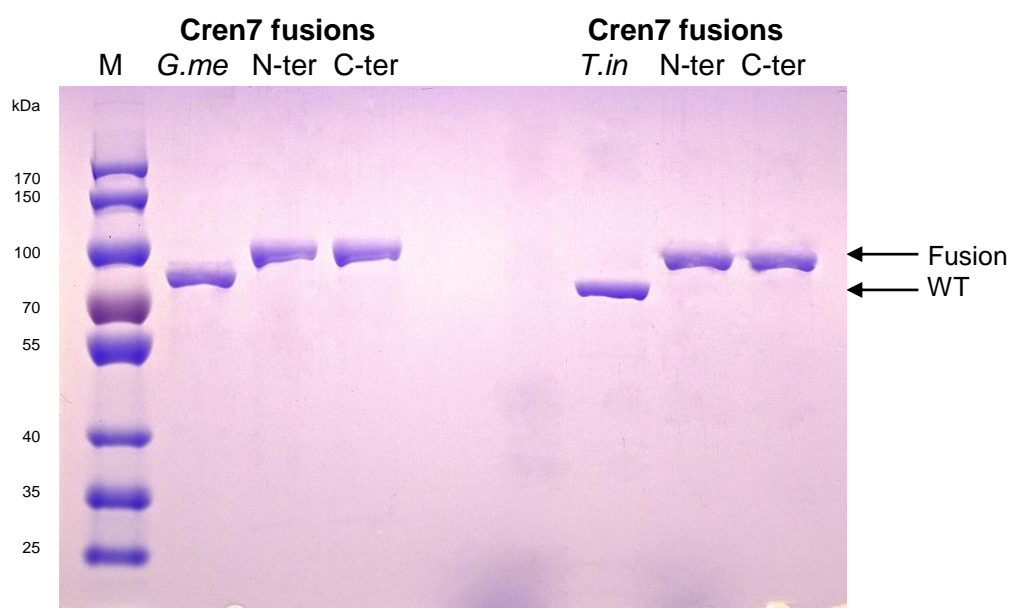
### *P.is* Cren7-*T.in* LF(exo-) LF DNA polymerase I

The same overlapping primer method was used to generate fusion proteins with *T.in*(exo-) Pol I. A summary of the overlap extension PCR, and the primers required, are detailed in Figure 8.3.



Purified fusion PCR products were digested with *Nde* I and *Bam*H I restriction enzymes for directional cloning into pET24a+HIS. Ligated DNA was transformed into *E.coli* KRX(pRARE2) and plated onto LB agar (*Kan*). Five colonies of each Pol I fusion were screened using T7 primers and the recombinant clones were DNA sequenced using the T7promoter or T7terminator primers to confirm the fusion. A single clone of each fusion was over-expressed on a large-scale, and purified as previously reported. 1ug protein was visualised on a 12% SDS-PAGE to confirm expression and purity (Figure 8.4).





**Figure 8.4**

1µg purified Cren7-Pol I fusion proteins on a 12% SDS-PAGE gel. M: PageRuler Marker, *G.me*: *G.me*\* Pol I I, N-ter: *P.isCren7-G.me*\* Pol I, C-ter: *G.me*\*-*P.isCren7* Pol I, M2: PageRuler PLUS marker, *T.in*: *T.in(exo-)* Pol I, N-ter: *P.isCren7-T.in(exo-)* Pol I, C-ter: *T.in(exo-)-P.isCren7* Pol I. Arrows identify the overexpressed Pol I (WT) and cren7-Pol I fusion proteins (Fusion). Arrows indicate the WT and the fusion Pol I protein bands.

## 8.2.2 Characterisation

The purified fusion proteins were compared to the wild-type enzymes in a number of assays to characterise activity.

### 5'-3' DNA polymerase primer extension

To ensure the fusion proteins were active, and to confirm the dsDNA binding protein or the linker region were not interfering with 5'-3' DNA polymerase activity, the enzymes were tested in the M13mp18 ssDNA primer extension assay as per Chapter 2.6.10.

The activity of *G.me*\* and *T.in(exo-)* Pol I were shown to be reduced when bound to the fusion protein (Table 8.1). A significant reduction was observed when the *P.is* Cren7 was fused to the C-terminus of *G.me*\* and *T.in(exo-)* Pol I with a 53- and 150-fold reduction, respectively.

The N-terminal fusion showed the least effect on the ability to extend the M13ssDNA bound primer. The N-terminal fusion to *G.me\** Pol I showed a 1.6-fold reduction in activity, and *T.in(exo-)* Pol I showed a 1.5-fold decrease. These results agree with the data from Wang *et al.* (2004) who only reported a fusion of Sso7d dsDNA binding protein to the N-terminal region of *T.aq* polymerase.

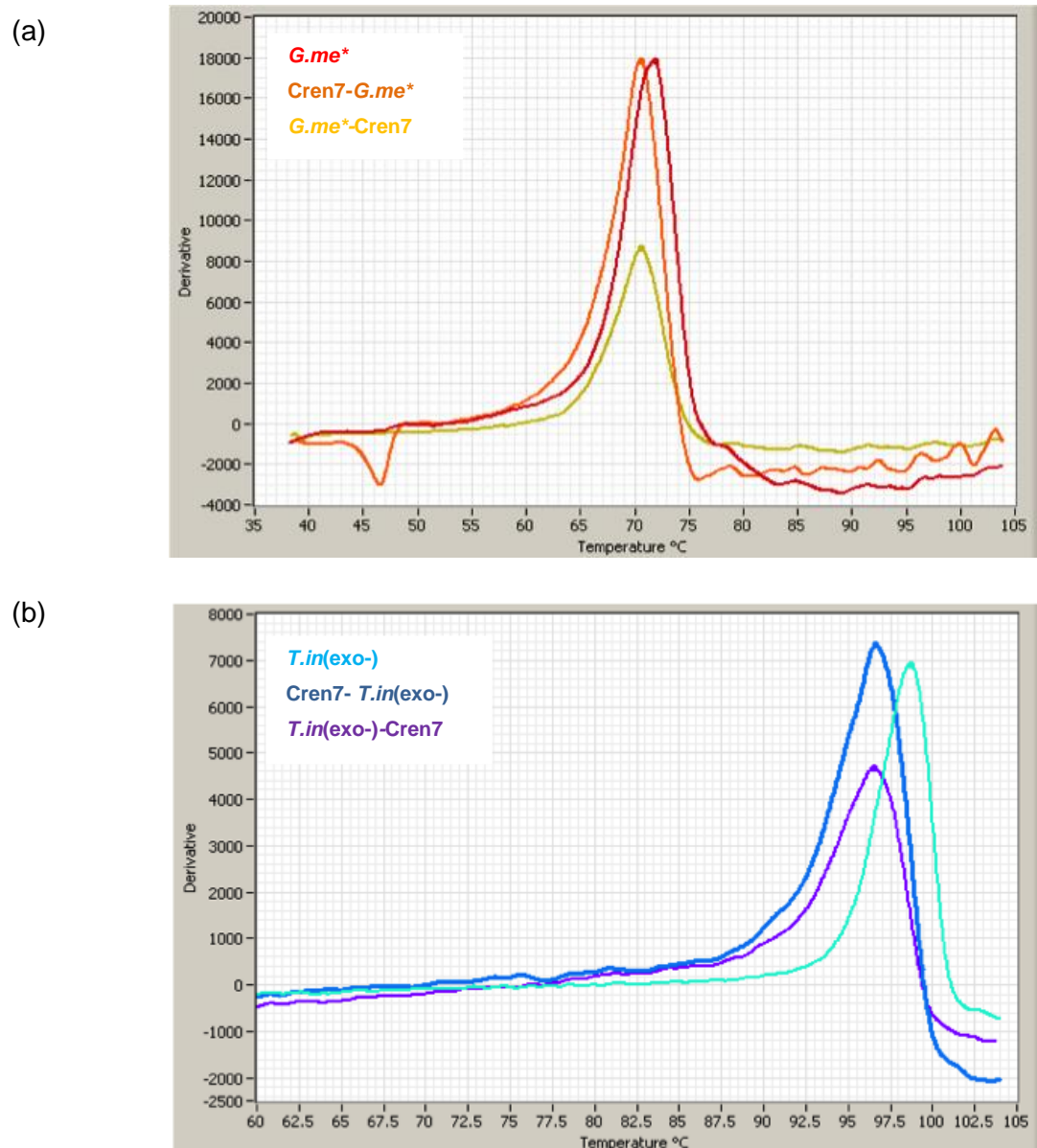
	Specific activity (Bst-equivalent units/mg)
<i>G.me*</i>	120,000
<i>C7-G.me*</i>	75,000
<i>G.me*-C7</i>	225
<i>T.in(exo-)</i>	7500
<i>C7-T.in(exo-)</i>	5000
<i>T.in(exo-)-C7</i>	50

**Table 8.1**

*A comparison of the Cren7 fusion proteins to the wild-type enzymes using the 5'-3' DNA extension assay to calculate B.st-equivalent unit.*

### Thermal melt analysis

Having first confirmed the fusion proteins were active, the effect of the addition of the DNA binding protein on thermostability was then tested. The fusion of binding proteins has been previously reported to increase the stability of nucleic acid modifying enzymes including, DNA Polymerase I (Pavlov *et al.*, 2002). 1µg of each Pol I was denatured in the presence of SYPRO®-Orange to compare the thermal melt profile with and without the cren7 fusion protein (Figure 8.5). The reaction was run as reported in Chapter 2.6.3.

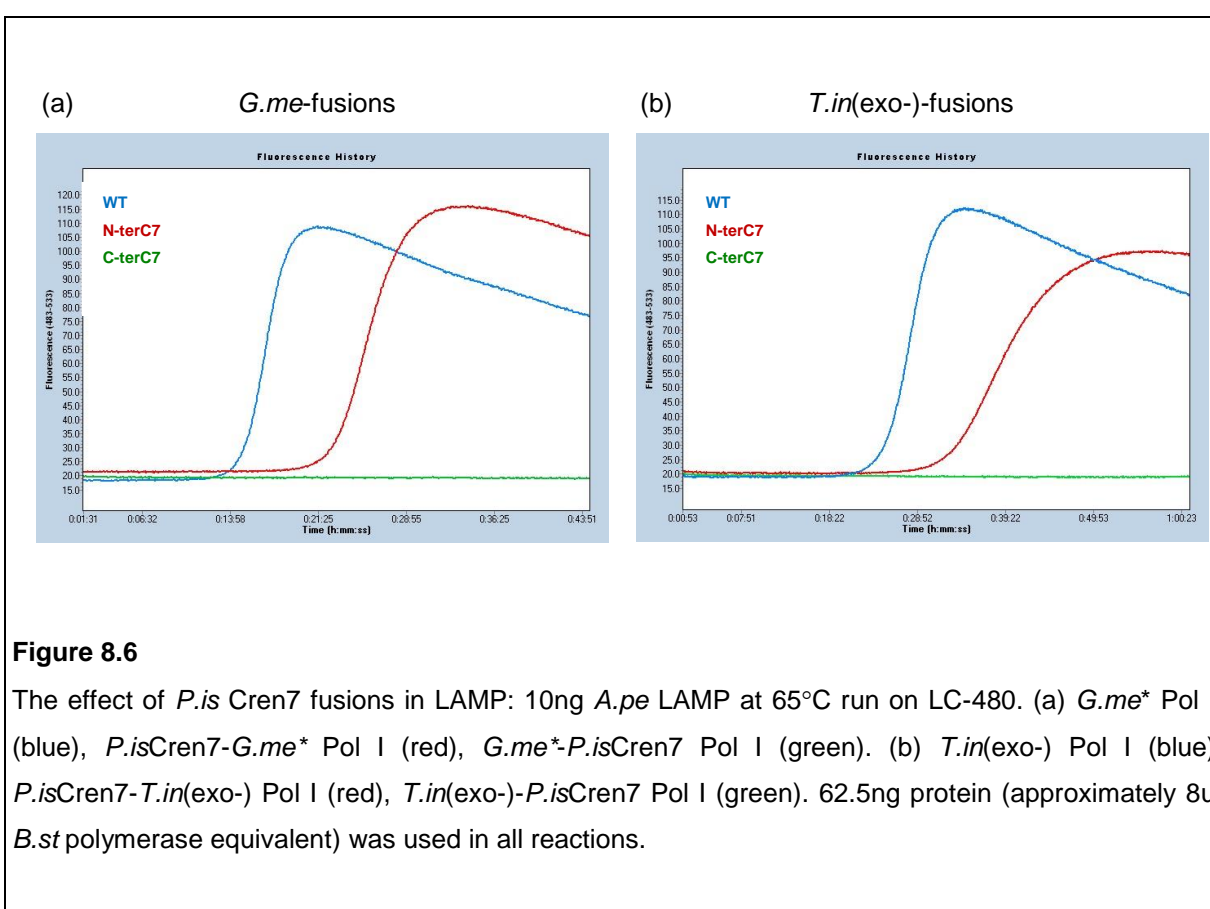


**Figure 8.5**

Thermal denaturation profiles comparing the effects of fusion proteins. (a) *G.me\** Pol I (red) , *P.isCren7-G.me\** Pol I (orange), *G.me\*-P.isCren7* Pol I (yellow), (b) *T.in(exo-)* Pol I (light blue), *C7-T.in(exo-)* Pol I (blue), *T.in(exo-)-P.isCren7* (purple).

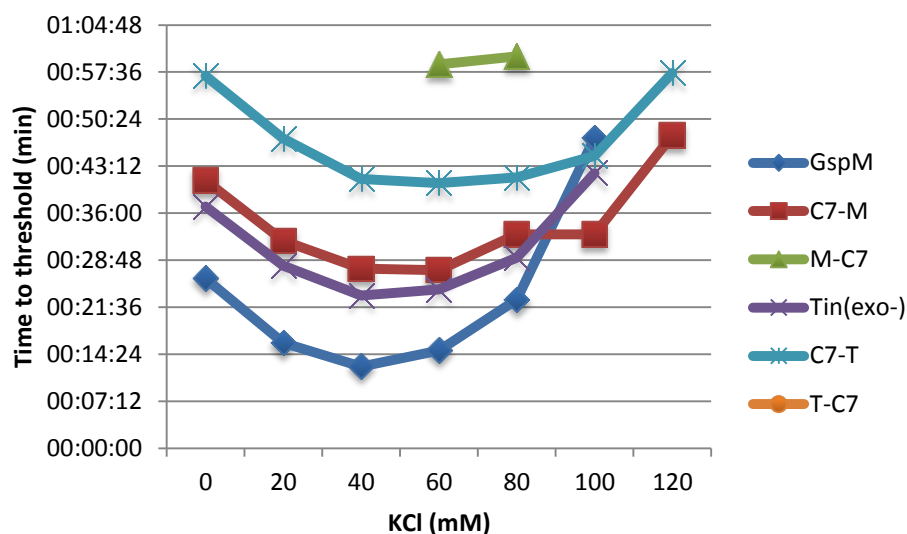
The fusion proteins were not shown to stabilise the polymerase with increasing temperature. The wild-type enzymes appear to be marginally more thermostable than the fusion constructs.

All polymerase fusions were shown to retain the thermostability required to tolerate the 65°C LAMP reaction. A dilution of each Pol I was used to amplify 10ng *A.pe* gDNA to compare the effect of the fusion on the ability to strand-displace in LAMP, following the method in Chapter 2.6.5. The results from Figure 8.6 clearly show the C-terminal fusion protein is not suitable for LAMP using either *G.me*<sup>\*</sup> or *T.in(exo-)* Pol I. Amplification was recorded >1h and therefore the strand-displacement activity is still present. It is expected this result is due to the significant reduction in extension activity of the polymerase rather than a loss of strand-displacement activity. The N-terminal fusion proteins showed positive results, with amplification.



### Salt tolerance

It was proposed that a DNA binding domain may confer resistance to an increased concentration of salt, as observed in a number of papers (Pavlov *et al.*, 2002; Wang *et al.*, 2004). To test this theory the LAMP reaction was repeated in the presence of increasing KCl concentrations from 0-120mM KCl (final) and the results are reported in Figure 8.7.



**Figure 8.7**

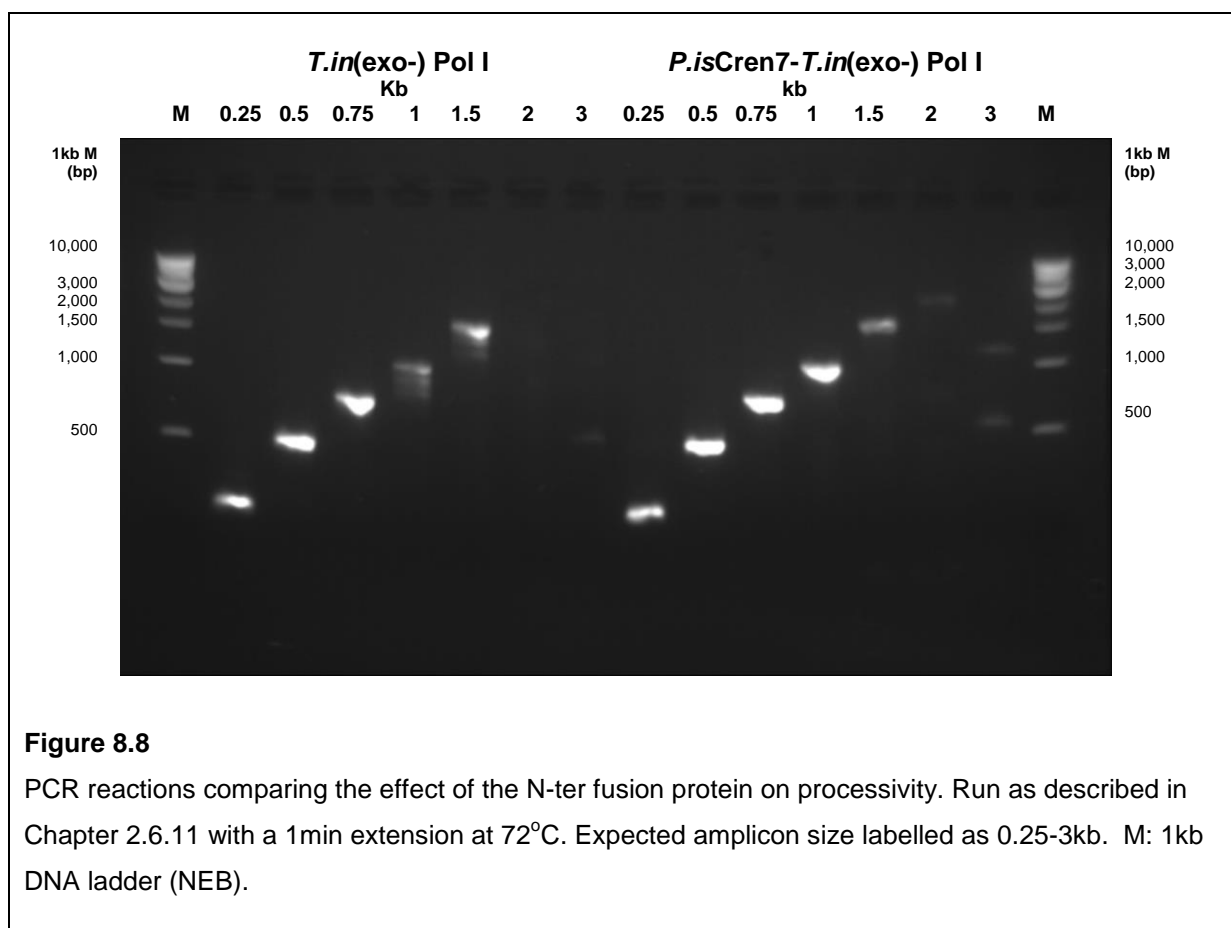
The effect of KCl concentration in LAMP, comparing wild-type and fusion proteins. Reactions were run following the method in 2.6.5 at 65°C. G: *G.me*, T: *T.in*, C7: *P.isCren7*. 62.5ng protein (approximately 8u *B.st* polymerase equivalent) used in all reactions.

The N-terminal fusion proteins *P.isCren7-G.me\** and *P.isCren7-T.in(exo-)* Pol I both showed a slight increase in the ability to perform LAMP in the presence of increased KCl. The wild-type enzymes did not amplify in the presence of 120mM KCl within the 1h reaction time. This result indicates the fusion protein may be a suitable alternative where reactions require amplification in the presence of high salt concentrations.

### Polymerase chain reaction

A further potential enhancement of the DNA binding protein that had not yet been explored was the effect on the processivity of the polymerase. The fluorescent M13 ssDNA extension assay had reported a reduction in DNA extension activity in LAMP, but the effects in PCR were unknown. The thermal melt analysis of the fusion proteins earlier in the chapter confirmed the *T.in* Pol I fusions retained their high thermostability, with only a marginal reduction observed. This implied the Pol I enzymes may be suitable for PCR, requiring high thermal denaturing steps for cycling amplification.

Reactions were run as reported in Chapter 2.6.11, using 66ng Pol I in each PCR and primers targeting 250bp-3kb Lambda DNA (Appendix i.vii). The extension time was fixed to compare the ability of the wild-type and fusion protein to amplify each fragment size within 1min. 5ul of the reaction product was analysed on a 1% (w/v) agarose gel (Figure 8.8).



The *T.in(exo-)* Pol I can be seen to amplify up to 1.5kb with a 1min extension time. Multiple bands appear to be visible in amplicons greater than 1kb. The addition of the cren7 binding domain can be seen to enhance the processivity of the *T.in(exo-)* Pol I enzyme, amplifying the 2kb fragment within the 1min extension time, but also reducing non specific amplification in the bands generated up to 2kb. Although outside the direction of this thesis, *P.isCren7-T.in* Pol I is suggested here as a suitable enzyme for specific PCR.

### 8.3 Discussion

This chapter has explored the use of Cren7, a double-stranded DNA binding protein, and its effects in loop-mediated isothermal amplification reactions.

The Crenarchaeon *Pyrobaculum islandicum* (*P.is*) is a hyperthermophile isolated by Stetter and Huber (1987). The organism has a reported growth temperature up to 100°C and one of its many archaeal chromatin binding proteins, cren7, has been demonstrated to enhance the performance of DNA polymerase enzymes in PCR (Clark *et al.*, 2009).

The use of DNA binding protein fusions has not previously been reported for use in loop-mediated isothermal DNA amplification reactions. The potential enhancements to LAMP were therefore unknown. *P.is* Cren7 was fused to the N- and C-terminal domains of *G.me*\* and *T.in*(exo-) Pol I, using a 3 amino acid GTH linker to separate the two proteins. It was hoped the fusion might confer the increased thermostability and activity observed previously in PCR (Clark *et al.*, 2009).

All fusion proteins demonstrated a reduced rate of dNTP incorporation compared to the wild-type enzyme. However, the N-terminal fusion was significantly more active than that of the C-terminal. This was observed with both polymerases. Wang *et al.* (2004) reported the fusion of Sso7d exclusively to the N-terminal of the Family A *T.aq* polymerase and to the C-terminal of the Family B *P.fu* polymerase. The results in this chapter further demonstrate a preference for N-terminal fusions, and this is expected to be a generalised result for all Family A polymerases.

It is reasoned the binding protein is interfering with DNA polymerase I activity when bound to the C-terminus due to its physical location and proximity to the catalytic Palm subdomain. *P.is* Cren7 may therefore interact with the template and or primer strand during nucleotide incorporation. The N-terminus of the polymerase, as observed from the crystal structures 3BDP (*B.st*) and 3KTQ (*T.aq*), is located beyond the 3'-5' exonuclease domain, far away from the catalytic active sites, therefore allowing free entry of the DNA primer-template and is unlikely to inhibit conformational changes of the polymerase during incorporation.

The N-terminal domains were further demonstrated to be active in LAMP but amplification activity was reduced when compared to wild-type enzymes. Due to the localisation of the N-terminal fusion proteins being a significant distance from the Fingers subdomain, it was

reasoned that the reduction in LAMP activity was not a loss of strand-displacement activity directly, but the effect of the reduced DNA incorporation activity.

Similar dsDNA binding protein fusions to thermophilic DNA polymerases report increased incorporation and processivity rates over the wild-type enzymes for use in PCR (Wang et al., 2004; Clark, 2009). The DNA binding proteins increase the affinity to the DNA template at the higher reaction temperatures required for PCR but this affinity is possibly not as important at the lower 65°C reaction temperature required for LAMP.

The results of inhibition to the LAMP activity observed during this study are supported by experiments from Huang *et al.* (2013). They investigated the role of *Sulfolobus* chromatin proteins Sso7d and S.so Cren7 in strand-displacement by DNA polymerase B1 from the hyperthermophile *Sulfolobus solfataricus* (S.so). DNA Polymerase B1 is believed to be the only replicative DNA polymerase in the hyperthermophile S.so and operates on both leading and lagging strand synthesis (Rogozin *et al.*, 2008), in conjunction with a variety of accessory proteins (Bell, 2012). Interestingly, structural analysis showed PolB1 to possess an unusual Fingers subdomain composed of four alpha helices (Savino *et al.*, 2004). As mentioned in Chapter 1, several helices within the Fingers domain are known to play key roles in DNA separation (Yuan *et al.*, 1998) and thus PolB1 may be structurally adapted for its robust strand-displacement activity (Huang, 2004). However, the requirement for several accessory proteins prevents PolB1 from being used *in vitro*.

Huang *et al.* (2013) demonstrated the DNA binding proteins Sul7d and Cren7 to modulate the strand-displacement activity of PolB1, reducing the size of displaced DNA to 3-4nt. The restraint imposed by the binding proteins is believed to enable the RNaseH to efficiently cleave the RNA primer during DNA repair. It is reasoned that the complex dsDNA structure of LAMP reactions may inhibit binding of the cren7 domain to the DNA template. Additionally, the cren7, as demonstrated by Huang *et al.* (2013) may further restrict activity at the site of strand-displacement. With the LAMP reaction relying on strand separation, which is not the case for PCR, the addition of binding proteins appears to be potentially inhibitory to amplification speed of LAMP.

One interesting result, confirming investigations by Wang et al. (2004) and Pavlov et al. (2008), demonstrated the ability to amplify DNA in the presence of increased levels of salt. N-terminal fusions were shown to amplify LAMP reactions in the presence of 120mM KCl within the 1 hour reaction. This is compared to the wild-type enzymes that amplified in a maximum concentration of 100mM KCl within the 1 h reaction. This is reasoned to be due



to the strong DNA template binding affinity of the cren7 protein, preventing the enzyme from uncoupling from the DNA, permitting activity in the higher salt concentrations.

The fusions were created using the short GTH linker, previously described and shown to produce active fusion proteins (Wang *et al.* 2004). Variations of this linker are further expected to alter the activity of the fusion proteins. Pavlov *et al.* (2008) reported the use of a long and flexible linker to fuse a TopoV domain to *T.aq* polymerase. They concluded the flexible linker enabled the fusion domain to physically interact and bind the enzyme, increasing the thermostability of the *T.aq* enzyme. Although not suitable for this study, the report suggests alternative linkers can be achieved and may offer further potential enhancements.

This chapter reports the successful fusion of dsDNA binding proteins to novel DNA polymerases. Although the results did not demonstrate positive enhancements to amplification speed, the ability to amplify in environments with increased levels of salts may be beneficial in a variety of applications. The LAMP reaction is reported as a portable alternative to PCR and so the ability to obtain pure template extraction may be limited in the field.

## 9. Final Discussion and Future Aspects

### 9.1 Thesis summary

The objective of the research reported in this thesis was to identify new DNA polymerases suitable for use in isothermal DNA amplification reactions. It was envisaged that a wide selection of suitable enzymes would increase our understanding of the unique mechanisms required by the polymerases for their application in loop-mediated isothermal amplification (LAMP).

#### Isolation and characterisation of novel DNA polymerases

Chapter 3 detailed the successful identification of twelve novel *polA* genes from a selection of thermophilic organisms using a variety of methods. The gene sequences were compared to those encoding well-documented Large Fragment DNA Pol I enzymes, including *E.coli* 'Klenow' and *Thermus aquaticus* 'KlenTaq', enabling the design of sequence-specific primers for directional cloning. Recombinant proteins were over-expressed and purified to apparent homogeneity in Chapter 4, enabling their characterisation and comparison.

Characterisation assays, in Chapter 5, identified the majority of DNA polymerases to demonstrate strand-displacement activity. However, not all were shown to be suitable for LAMP. The *Geobacillus*-like DNA polymerases showed the fastest amplification time in LAMP. This was not an unexpected result, with the current gold-standard LAMP enzyme being the commercially-available *B.st* LF DNA Pol I from *Geobacillus stearothermophilus* (NEB, UK). The *Geobacillus* Pol I enzymes reported in this work could therefore be used as direct alternatives to *B.st* LF DNA Pol I. However, no further novel applications over the existing commercially available enzyme were identified using the *Geobacillus*-like Pols in this thesis. The remaining strand-displacing DNA polymerases did not amplify as fast in LAMP. Several enzymes, however, were shown to be substantially more thermostable than the *B.st* LF DNA Pol I enzyme.

Thermal melt analysis proved to be a fast and accurate technique to monitor the real-time denaturation of proteins over a temperature gradient. In combination with the Genie<sup>®</sup>II, enabling highly sensitive real-time fluorescence detection, the SYPRO-Orange<sup>®</sup> protein binding dye gave specific and highly reproducible thermal denaturation characteristics for each enzyme under test. Previous reports using real-time thermal melt analysis have relied

on the use of qPCR thermal cycling instrumentation, with the integrated analysis software not specifically designed for the precision gradient analysis required for highly accurate results to be obtained. The Genie<sup>®</sup>II is therefore identified in this thesis as the instrument of choice for protein melt analysis.

The most novel polymerases were found to be the DNA Pol I enzymes from the *Thermodesulfatator* species. These polymerases showed ~50% amino acid sequence identity to the closest match currently in the public database. *Thermodesulfatator indicus* (*T.in* Pol I) and *Thermodesulfatator atlanticus* (*T.at* Pol I) demonstrated strong strand-displacement activity at high reaction temperatures. The polymerases were shown to be highly thermostable, well above the maximum reported growth temperature for this species. *Thermodesulfatator* sp. were isolated from deep-sea hydrothermal vents (Moussard *et al.*, 2004; Alain *et al.*, 2011). These superheated sites are dominated by highly thermostable organisms. There are limited reported data on the *Thermodesulfatator* species. Questions arise that ask why two *Thermodesulfatator* species encode such highly thermostable polymerases, given their optimum reported growth temperatures of ~70°C. With limited available information as to the exact ecosystem in which they are found, it is difficult to draw conclusions to the exact growth conditions of these species. It is reasonable to expect that all organisms require their replicative enzymes to survive conditions above the upper limit of their growth, to enable cell division and survival at the extremes of their limits. With conditions at and around a hydrothermal vent varying from psychrophilic to hyperthermophilic, depending on the distance from the vent, it is possible horizontal gene transfer could have occurred between *Thermodesulfatator* species and a more hyperthermophilic organism. It is also possible the *Thermodesulfator* species have evolved from a more thermostable organisms living closer to the vent. Life at a hydrothermal vent is complex and it is therefore difficult to draw exact conclusions.

The genome sequence of *Thermodesulfatator indicus* has recently been reported (Anderson *et al.*, 2012), confirming the sequence of the *T.in* Pol I in this report. The full sequence of *Thermodesulfatator atlanticus* is also close to completion. Examination of these genomes, and a comparison with those from alternative hydrothermal vent species, may highlight these genetic transfer events.

Having identified alternative LAMP enzymes, it was important to investigate the specific reaction buffers for each enzyme. Work in Chapter 5 ensured each enzyme was suitably active, enabling accurate comparisons. All polymerases showed a relatively wide tolerance to individual buffer components and the commercially available LAMP reaction buffer chosen

was shown to be an effective buffer for all enzymes studied. Precise buffer optimisation may further enhance individual LAMP reactions, but not significantly from the standard reaction conditions reported in this study. Additional optimisation may be required if, for example, the desired target varies greatly in the GC content from the *Aeropyrum pernix* (*A.pe*) target used for the standard characterisation assays in this thesis.

The varying ability of Family A DNA polymerases to strand-displace, highlighted in Chapter 5, raises the fundamental question of why some Pol I enzymes are observed to strand-displace more efficiently than others? This is especially intriguing when all Family A DNA Pol I enzymes are required for the same function *in vivo*, namely the RNA primer removal of Okazaki fragments during lagging-strand synthesis. Results from Chapter 5 led to the observation that an organism's reported optimal growth temperature relates to the ability of its DNA Pol I to strand-displace. Mesophilic DNA polymerases (Family A *E.coli* Klenow DNA Pol and Family B *Bacillus* phage Phi29 DNA Pol) are reported to have extremely high strand-displacement activity. The thermophilic strains in this study, including the *Geobacillus*-like species, further highlight this correlation to increasing growth temperature with a reduced ability to strand-displace. It is reasoned that the more thermophilic organisms may require accessory proteins to enable sufficient strand-displacement *in vivo*. As the growth temperature increases, so the DNA denaturation, the breaking of hydrogen bonds, is energetically more favourable at the higher temperatures. This denaturation is expected to be increasingly high in the absence of double-stranded DNA binding proteins, required for DNA stability at increased temperatures. Such proteins include SSB (Marceau, 2012), Cren7 (Huang *et al.*, 2008).

It could therefore be reasoned that mesophilic organisms require strong strand-displacement activity to enable efficient DNA repair at the less energetically favourable DNA denaturation temperatures. In contrast, as the environmental temperature increases, so the requirement for strand-displacement activity decreases, due to the DNA being more easily separated. When a replicative polymerase stalls at a site of DNA damage, DNA repair requires the localisation of accessory proteins, including DNA-binding proteins, to the site of strand separation. These binding proteins can restrict the energetically favourable strand-separation of DNA. This observation appears to hold true with the polymerases identified in this thesis: organisms isolated from moderately thermophilic conditions (60-70°C) show the fastest rate of strand-displacement activity in this report.

It must be noted that the reaction temperature in LAMP is defined by the primers and the enzyme in use, and is further restricted by the melting temperature of the DNA amplicon

itself. The polymerases identified in this thesis showed optimal strand-displacement activity at 65°C. The assay temperatures within this study therefore dictate the observed characteristics. The mesophilic *E.coli* Klenow and *Bacillus* phage Phi29 polymerases are reported to show significant strand-displacement activity but they are not sufficiently thermostable to be suitable for diagnostic amplification reactions (e.g. LAMP), where the increased temperatures provide the specific annealing of primers that are required to impart the high specificity to the reaction.

It is important to remember that the enzymes have been characterised in conditions often very different from their natural environment, and excluding all possible accessory proteins that may be available to the enzyme *in vivo*. For this reason, conclusions as to the exact role and activity of the enzyme *in vivo* cannot be determined. These novel enzymes have been compared in defined assay conditions favourable to the application to be improved. In this respect *T.in* Pol I is the most interesting enzyme identified, offering the strand-separation activity required for LAMP, yet uniquely thermostable to enable a novel high temperature step. The initial objective to identify novel DNA polymerases for isothermal amplification was therefore achieved in this report. The results in this thesis suggest future exploration of novel enzymes suitable for LAMP be directed at mesophilic to mildly thermophilic organisms that may encode Pol I enzymes with strong strand-displacement activity but will tolerate the isothermal temperatures required for a highly specific reaction.

### Polymerase engineering

There has been limited research into the exact nature of strand-displacement activity of DNA polymerases. The last 20-30 years of polymerase research has been driven by their application to PCR and sequencing in the biotechnology field. A wide variety of thermostable polymerases have been identified and further engineered for these applications, namely to modify the fidelity, processivity, and nucleotide incorporation ability of the enzyme.

With the relatively recent introduction of isothermal methods, interest must be drawn to the enzymes required, and this was the direction for further study in this thesis.

Strand-displacement activity has been reported for several Family A and B polymerases, yet little is currently understood about the exact nature of this activity. Strand-displacement is well documented to be required during lagging strand synthesis, but why some Pol I enzymes demonstrate superior activity over others *in vitro* has not been investigated.

The current understanding of Pol I strand-displacement activity is directed towards the Fingers subdomain of the polymerase to physically separate the two DNA strands before the point of nucleotide incorporation in the Palm subdomain. This understanding has been further directed to three highly-conserved residues, with Singh *et al.* (2007) reporting the S F/Y R motif to be an absolute requirement of *E.coli* Klenow Pol I for strand separation. The group inferred that this motif is a requirement for all Pol I enzymes to possess strand-displacement activity.

Work in Chapter 6, however, showed this not to be the case. Site-directed mutagenesis of these key residues failed to impart identifiable strand-displacement activity to *T.aq* or *T.ma* Pol I. Furthermore, saturation mutagenesis of these key residues in *G.me*\* Pol I and *T.in*(exo-) Pol I showed the nature of the residues to be relatively flexible. The majority of mutated enzymes remained active in LAMP, and therefore they still retained the strand-displacement activity. A single residue (Arg<sup>789</sup> in *B.st* Pol I), however, was shown to be absolutely required for strand-displacement activity, confirming the observation reported by Singh *et al.* (2007) that the highly conserved arginine stabilises the template strand during DNA synthesis.

This flexibility, from two of the reported key strand-displacement residues, indicated that alternative residues and/or motifs must be required to impart the strand-displacement activity to the Pol I enzyme. Further work in this thesis was therefore directed at the identification of these residues with the hope of enhancing the novel enzymes for use in LAMP.

Results in Chapter 7 of this thesis confirm the location of strand-displacement activity to the Fingers subdomain. Joyce *et al.* (1995) first suggested the Fingers subdomain play a role in DNA binding of the template strand. To confirm this observation, enzyme chimeras were designed in this report to cover a variety of the Pol I subdomains, including the Fingers subdomain. In order to create the chimeras, known strand-displacing enzyme residues were swapped into a polymerase with no observable strand-displacement activity. This thesis details the exclusive use of the Fingers subdomain to impart strand-displacement activity to *T.aq* Pol I. The Fingers subdomain of *T.in* Pol I moved into *T.aq* Pol I is, to my knowledge, the first reported demonstration of a *Thermus* sp. Pol I enzyme suitable for use in a LAMP reaction. Truncated forms of *T.aq* Pol I have been reported to show strand-displacement activity (Guoliang, 2012) but no data can be found demonstrating their use in an isothermal DNA amplification method and, as such, cannot be verified.

Due to the large number of chimeras required for this study, the recombinant chimeric enzymes were not fully purified. It would be interesting to explore these purified chimeras further for a comparison of their thermostability. The results in Chapter 7 clearly demonstrate the importance of the Fingers subdomain in DNA binding and its role in strand displacement.

This thesis identified novel DNA polymerases and gained information as to the nature of their strand-displacement activity. Chapter 6 demonstrated the importance of reducing the 3'-5' exonuclease activity of the Pol I for use in LAMP in that this activity was shown to degrade the oligonucleotide ssDNA primers. Site-directed mutagenesis of key residues reduced the observed 3'-5' exonuclease activity, and such mutant enzymes were shown to amplify faster in LAMP, when compared to the wild-type enzymes, and were therefore used for all further studies.

Chapter 8 attempted to further enhance the novel enzymes with the addition of a double-stranded DNA binding protein. Cren7 has been shown to enhance the processivity and salt tolerance of Pol I (*T.aq*) and II (*P.fu*) in PCR (Clark *et al.* 2009). Although using a different DNA amplification method, it was hoped a similar enhancement could be observed in LAMP. Cren7 was fused to either the N- or C-terminal regions of *G.me*\* Pol I and *T.in(exo-)* Pol I. In this thesis, Cren7 fusions were shown to be detrimental to the LAMP reaction by significantly increasing the time for amplification. However, one advantage of the fusion proteins was to enhance polymerase activity in the presence of increased salt concentrations.

## **9.2 Industrial applications**

The objective for this thesis was to identify new DNA polymerases for isothermal amplification. The number of applications for isothermal nucleic-acid amplification is increasing as the need for on-site detection of diseases and pathogens is increasingly desired. This further highlights the limitations to current nucleic-acid based detection systems such as PCR.

The main aim of deploying detection methods at the point-of-sampling, or point-of-care, is to enable evidence based decisions to be taken more rapidly. The ability to run on-site nucleic acid based testing further re-inforces the credibility of those decisions. This will be especially important with new and emerging disease threats where outbreaks must be diagnosed, and individuals treated or quarantined, as soon as practically possible. A further

advantage of nucleic acid-based detection is the ability to rapidly develop specific primers to detect new sequence targets. Next generation sequencing technology further enables the fast identification of an organism's genome, enabling the design of unique target primers.

Novel polymerases may lead to potentially novel applications and enhancements over the existing methods. *T.in(exo-)* Pol I, identified in this thesis, is one such enzyme offering new applications for nucleic-acid amplification, not currently achievable using the gold-standard *B.st* Pol I.

### HD-LAMP

Heat-denaturing LAMP (HD-LAMP), detailed in Chapter 5, is the first reported method allowing single-step, closed-tube template denaturation with target detection in LAMP. The original LAMP paper (Notomi *et al.*, 2000) suggests the use of an initial 95°C, 5-minute heat step, followed by chilling on ice to allow the addition of *B.st* Pol I. Although not an absolute requirement of LAMP (Nagamine *et al.*, 2001), this heat step has been reported to increase analytical sensitivity (Suzuki *et al.*, 2010; Aryan *et al.*, 2010; Geojith *et al.*, 2011). Chapter 5 highlighted a further advantage, that of reducing amplification times, when an initial heat-denaturation step was included. This HD-LAMP step is uniquely available to *T.in(exo-)* Pol I. A closed-tube reaction held at 95°C for 5 min enables applications not previously achievable in a closed-tube reaction. This initial, pre-LAMP heat step can be envisaged to aid sample lysis, denaturing target organisms, enabling the direct detection of genomic DNA with no DNA extraction step.

The enhancement of LAMP using an initial template denaturation step raised an intriguing question as to how isothermal amplification is initiated in LAMP? To enable a successful LAMP reaction, ssDNA primers must anneal to dsDNA. Without an initial heat-step, absolutely required for PCR, isothermal reactions require alternative methods to separate the dsDNA strands. LAMP primer annealing provides multiple strand separation points (reviewed in Chapter 1), but how does the first primer anneal to the dsDNA enabling the initial dumbbell-shaped product formation to occur?

It has been suggested that the 65°C reaction temperature enables the spontaneous formation of local denaturation bubbles in the Watson-Crick double strands, providing localised sections of single-stranded DNA (Kornberg, 1978, and Metzler, 2009). When denaturation bubbles open up at the chosen priming sites of DNA in a LAMP reaction mixture at 65°C, the formation of a primer-template hybrid is favoured over re-annealing of



the template strands by the high ratio of the concentration of primer to template DNA and the inclusion of betaine in the reaction (Wartell *et al.*, 1985). Once formed, the primer within the primer-template hybrid is elongated by *B.st* Pol I. The extension and strand-displacement of any upstream primer opens the double-stranded target DNA and exposes the binding sites for the downstream primer (Xu *et al.*, 2012). After the initial amplification step, LAMP relies upon the formation of a dumbbell structure encoded within the 5' tail of the primers (Nagamine *et al.*, 2001).

However, this theory of isothermal amplification initiation is contradicted by Mitsunaga *et al.* (2013). The authors showed the pre-treatment of DNA samples with T4 DNA ligase repaired all nicks in the DNA. They showed DNase I treated DNA to amplify in LAMP but no amplification was observed in the undigested sample. This leads to the suggestion that the strand-displacement activity of DNA polymerase in LAMP, at least in part, starts from randomly existing nicks. In this case, the time required for the strand-displacing enzyme to displace from the nick to the primer/template site will be a limiting step of the LAMP reaction. HD-LAMP will therefore avoid this rate-limiting step through the template denaturation allowing efficient primer annealing at the 65°C reaction temperature.

An isothermal nucleic acid amplification review (Craw *et al.*, 2012) concluded "the limiting factor in the nucleic acid point-of-care test (POCT) product development pipeline is not the availability of suitable techniques but the integration of existing amplification techniques with upstream sample processing and nucleic acid isolation methods with downstream detection schemes". The use of *T.in(exo-)* Pol I may offer one such solution.

The initial heat step, exclusively available to *T.in(exo-)* Pol I, is also thought to provide conditions favourable for highly-specific primer annealing. At 95°C the LAMP primers cannot anneal to the target DNA or form primer dimers. As the temperature reduces to the 65°C LAMP reaction temperature, the primers will anneal to the target DNA with greater specificity, reducing false-priming and thereby false amplification products observed with primer annealing at lower temperatures. This reduces the requirement for a hot-start enzyme, often desired in PCR, to reduce non-specific amplification during set-up.

HD-LAMP offers reduced amplification time, with increased sensitivity and specificity, all highly desired enhancements to a DNA diagnostic application, previously unobtainable within a closed-tube system. The closed-tube system significantly reduces chances of contamination by negating the need to open the reaction tube for the addition of the less thermostable *B.st* Pol I enzyme. Fewer manipulations prevent carry-over contamination with

amplification products and cross contamination between samples.

### Carry-over contamination

This thesis further details the importance of LAMP to be run as a closed tube reaction. Real-time monitoring and confirmation must be achieved without the need to open the reaction tube for end-point analysis. The reaction speed enables large quantities of amplicon to be generated and the contamination risks of opening the tube are substantial.

To further reduce the risk of amplicon contamination, the use of dUTP for carry-over protection could also be applied. Results in Chapter 5 showed *T.in* and the *Geobacillus*-like Pols to successfully incorporate dUTP. However, the dUTP-containing amplicon must be degraded by uracil DNA glycosylase (UNG) for effective carry-over protection. Additional work at GeneSys Biotech Ltd. has shown that *E.coli* UNG requires a high-temperature step for complete denaturation, which also degrades the *Geobacillus* Pol I, inactivating the enzyme. An alternative is to source a thermolabile UNG to enable complete removal of activity. The 65°C LAMP reaction temperature may prove too low to provide effective, and more importantly fast, degradation of the UNG. *T.in* Pol I, however, can survive this denaturation temperature, highlighting a further potential benefit of this novel enzyme. Thus, carry-over protection using *T.in* Pol I in HD-LAMP is suggested to offer significant advantages over existing *B.st* Pol I LAMP reactions and is an area for future investigation.

### Polymerase chain displacement reactions

The novel thermostable and strand-displacement properties further highlight *T.in*(exo-) Pol I for use in Polymerase Chain Displacement Reactions (PCDR). PCDR, which uses multiple nested primers in a rapid, single, closed-tube reaction, has been shown to increase the sensitivity of normal quantitative qPCR assays (Harris *et al.*, 2013). In PCDR, when extension occurs from the outer primer, it displaces the extension strand produced from the inner primer by utilizing a polymerase that has strand-displacement activity. This allows a greater than 2-fold increase of amplification product for each amplification cycle and therefore increased sensitivity and speed over conventional PCR. A modified *T.aq* DNA polymerase, reported to offer improved strand-displacement activity, was used by Harris *et al.* (2013) in that report. However, no reference was made to the exact modifications, and no data reported, or suggested, for its use in isothermal amplification reactions. Chapter 8 reported the ability of *T.in*(exo-) Pol I to be used in PCR and it is therefore suggested that it might be highly effective in the PCDR method utilising its strong strand-displacement activity.

### 9.3 Future aspects

One area of further investigation, only briefly covered in this study, is to employ additional accessory proteins that may be required to convey effective strand-displacement activity to the polymerase. This may include *T.aq* Pol I, which is reported to require single-stranded DNA binding proteins (SSB) *in vivo* (Viguear *et al.*, 2001). However, the nature of diagnostic assays favour single enzyme reactions. Buffer optimisation for *in vitro* reactions is complex and is particularly difficult with multiple enzymes, each requiring defined, and possibly differing, conditions for optimal activity. Furthermore, accessory proteins may require ATP and therefore ATP regeneration systems within the reaction may be needed, further complicating reaction components (Li *et al.*, 2008). The addition of thermophilic SSB has been reported to increase the specificity of LAMP reactions (patent application Tanner *et al.*, 2013). However, additional work conducted at GeneSys Ltd. has failed to confirm these enhancements.

The complex nature of multi-enzyme reactions within LAMP was not investigated in this thesis, but is envisaged to be an area for future assay development. Obviously, the individual enzyme components would need to be thermostable and not inhibit the fundamental nature of the strand-displacing DNA polymerase enzyme.

Unfortunately, crystallographic studies were not successful during this project. Current DNA Pol I structural data are provided from the mesophilic *E.coli* Klenow, and the moderately thermostable *B.st* and thermostable *T.aq* Pol I enzymes. Additional crystal data based on *T.in* Pol I may reveal more information on the exact nature of its high thermostability and strand-displacement activity, and will be an exciting area of future work.

Little is currently known regarding the interactions with the distant single-stranded template strand and the polymerase. Joyce *et al.* (2003) point-mutated possible contact residues on the Fingers subdomain. They first identified S<sup>769</sup>, F<sup>771</sup> and R<sup>841</sup> (*E.coli* Pol I numbering) to contact the ssDNA template strand. Their data suggested the template strand followed a path over the Fingers subdomain, close to a cluster of positively-charged residues. Ser<sup>769</sup> and Phe<sup>771</sup> are located on the O-helix of DNA pol I as defined by Ollis *et al.* (1985). The O-helix is made up of O, O1 and O2 helices. This 3-helix bundle has been shown to share structural homology with the DNA-binding motif Mrf-2 (Yuan *et al.*, 1998). Singh *et al.* (2007) further suggested that the motif is a distinct functional unit that recognises or induces altered DNA structure and requires an aromatic residue at its apex for strand separation.

With the confirmation by Singh *et al.* (2007) that these residues are important for strand-displacement activity, and the data in this thesis confirming the importance of unique Fingers subdomain interactions enabling strand-displacement activity, it is suggested that tighter binding to the template strand may further enhance strand-displacement activity to the Pol I. This may be achieved through further site-directed mutagenesis of the Fingers residues, close to the contact with the template strand, where the introduction of positively-charged residues may enhance the binding to the negatively-charged phosphate backbone of the DNA. Which residues to target, and whether this will hold true, will require further exploration. Here, a *T.in* Pol I crystal structure, in complex with the ssDNA template overhang, would help locate those specific residues. A greater understanding of the Mrf2-DNA interactions, introduced in Chapter 7 and 8, may also highlight these potential target motifs.

The overall fidelity of LAMP is not thought to be an essential component of the reaction in this thesis because the detection is a yes/no event, i.e the amplified product is not required for downstream applications such as cloning. However, a reduction in fidelity might reduce amplification speed and efficiency. The 3'-5' exonuclease active polymerases are expected to show a higher fidelity over those without an active proof-reading domain. Here the use of phosphothioated LAMP primers should be used throughout to avoid primer degradation during the reaction, as reported in Chapter 5.

One activity not identified in this thesis is the ability of a strand-displacement polymerase to efficiently copy RNA i.e. to show RNA-dependent DNA polymerase activity. Current RT-LAMP methods require a two-step approach, whereby a specific reverse transcriptase enzyme, for example Avian Myeloblastosis Virus Reverse Transcriptase (AMV-RT), is added prior to 65°C LAMP amplification. Engineering of the Pol I Palm subdomain, including the dNTP binding pocket, may enable the efficient incorporation of the bulky OH-group present in RNA. A highly-thermostable, strand-displacing RNA and DNA-dependent polymerase would be highly desirable to enable faster RNA and viral pathogen detection.

#### Importance of isothermal amplification

Isothermal methods are now enabling the sensitive and specific detection of nucleic acids at the point of sampling. They only require small, cheap, portable instrumentation, thereby increasing accessibility to scientists in developing countries. Isothermal reactions provide extremely rapid amplification that is now only restricted by the initial sample preparation and purification. This must now be the area for future development, to incorporate sample

preparation, target amplification and detection within a closed system. High temperature denaturation may be one such DNA extraction method, using the novel *T.in(exo-)* Pol I reported in this study, but it may not be suitable for all targets and templates.

PCR still has its place in nucleic acid testing. It is a powerful technique, but it is restricted to the laboratory due to the large, power-demanding instrumentation required to facilitate the cycling nature of the reactions. Furthermore, the cycling nature of PCR increases reaction time, reducing the suitability for a point-of-care reaction. LAMP can be envisaged to be a first-stage screening process, quickly identifying diseases and pathogens, directing further, more complex analysis, if required, back at the laboratory. The key advantage of isothermal detection is the ability to run assays with the same sensitivity and specificity to PCR at the point-of-sampling, be it in the field, or at the point-of-care. The speed of the reaction enables faster diagnosis and therefore faster treatment. Where there is still a requirement for final diagnosis to be confirmed within the laboratory using a legislated method, on-site detection for initial screening by LAMP would increase the credibility of the initially suspected sample.

In conclusion, the novel enzymes in this thesis further increase the suitability of LAMP as a means of fast, sensitive and specific DNA detection and offer the potential for a variety of new applications.

## References

- Alain** K, Postec A, Grinsard E, Lesongeur F, Prieur D, Godfroy A. 2010. *Thermodesulfatator atlanticus* sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a Mid-Atlantic Ridge hydrothermal vent. *Int J Syst Evol Microbiol.* **60**(1):33-38.
- Aliotta** J.M, Pelletier J.J, Ware J.L, Moran L.S, Benner J.S, Kong H. 1996. Thermostable *Bst* DNA polymerase I lacks a 3'-5' proofreading exonuclease activity. *Genet Anal.* **12**(5):185-195.
- Aliotta** J.M, Kong H. Pelletier. 1998. Expression product is substantially free of 3' to 5' exonuclease. Patent: US5814506A.
- Altschul** S.F, Gish W, Miller W, Myers E.W, Lipman D.J. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403-410.
- Anderson** I, Saunders E, Lapidus A, Nolan M, Lucas S, Tice H, Del Rio T.G, Cheng J.F, Han C, Tapia R, Goodwin LA, Pitluck S, Liolios K, Mavromatis K, Pagani I, Ivanova N, Mikhailova N, Pati A, Chen A, Palaniappan K, Land M, Hauser L, Jeffries C.D, Chang Y.J, Brambilla E.M, Rohde M, Spring S, Göker M, Detter J.C, Woyke T, Bristow J, Eisen J.A, Markowitz V, Hugenholtz P, Kyrpides N.C, Klenk H.P. 2012. Complete genome sequence of the thermophilic sulfate-reducing ocean bacterium *Thermodesulfatator indicus* type strain (CIR29812(T)). *Stand Genomic Sci.* **6**(2):155-64.
- Angerer** B, Ankenbauer W, Bonch-Osmolovskaya E, Ebenbichler C, Laue F, Markau U, Reiser A, Schmitz-Agheguian G, Svetlichny V. 1998. Thermostable DNA polymerase from *Carboxydotherrmus hydrogenoformans*. Patent: WO1998014589A1.
- Arnold** K, Bordoli L., Kopp J, and Schwede T. 2006. The SWISS-MODEL Workspace: A web-based environment for protein structure homology modelling. *Bioinformatics.* **22**:195-201
- Aryan** E, Makvandi M, Farajzadeh A, Huygen K, Bifani P, Mousavi S.L, Fateh A, Jelodar A, Gouya M.M, Romano M.. 2010. A novel and more sensitive loop-mediated isothermal amplification assay targeting IS6110 for detection of *Mycobacterium tuberculosis* complex. *Microbiological Research.* **165**(3):211-220.
- Astatke** M, Ng K, Grindley N.D.F. and Joyce C.M. 1998. A single side chain prevents *E.coli* DNA Polymerase I (Klenow fragmnet) from incorporating ribonucleotides. *PNAS USA*, 95 3402-3407.
- Asiello** P.J. Baeumner A. J. 2011. Miniaturized isothermal nucleic acid amplification, a review. *Lab Chip.* **11**(8):1420-1430.

**Ausubel** F.M, Brent R, Kingston R.E, Moore D.D. Seidman J.G, Smith J.A, Struhl K. 1994. Current Protocols in Molecular Biology, pp 2.0.1-2.14.8, John Wiley and Sons, New York.

**Bedinger** P, Munn M, Alberts B.M. 1989. Sequence-specific pausing during *in vitro* DNA replication on double-stranded DNA templates. *J Biol Chem.* **264**:16880-16886.

**Beese** L.S, Friedman J.M, Steitz T.A. 1993. Crystal structures of the Klenow fragment of DNA polymerase I complexed with deoxynucleoside triphosphate and pyrophosphate. *Biochemistry.* **32**(51):14095-14101.

**Kiefer** J.R, Mao C, Braman J.C, Beese L.S. 1998. Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature.* **391**:304-307 .

**Beattie** T.R, Bell S.D. 2012. Coordination of multiple enzyme activities by a single PCNA in archaeal Okazaki fragment maturation. *EMBO J.* **31**(6):1556–1567.

**Bentsink** L, Leone G.O, van Beckhoven J.R, van Schijndel H.B, van Gemen B, van der Wolf J.M. 2002. Amplification of RNA by NASBA allows direct detection of viable cells of *Ralstonia solanacearum* in potato. *Journal of Applied Microbiology.* **93**:647-655.

**Bergquist** P.L, Shandilya H, Griffiths K, Flynn E.K, Astatke M, Shih P.J, Lee J.E, Gerard G.F, Gibbs M.D. 2004. Thermophilic bacterial DNA polymerases with reverse-transcriptase activity. *Extremophiles.* **8**(3):243-51.

**Blanco** L, Salas M. 1984. Characterization and purification of a phage phi29-encoded DNA polymerase required for the initiation of replication. *Proc Natl Acad Sci USA.* **81**(17):5325–5329.

**Bradford** M.M. 1976. Rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Anal. Biochem.* **72**:248–254

**Braithwaite** D.K, Ito J. 1993. Compilation, alignment, and phylogenetic relationships of DNA Polymerase. *Nucleic acids research.* **21**(4):787-802.

**Brock** T.D, Freeze H. 1969. *Thermus aquaticus* gen. n. and sp. n., a nonsporulating extreme thermophile. *J Bacteriol.* **98**(1):289-297.

**Çağlayan** M, Bilgin N. 2011. Cloning and Sequence Analysis of Novel DNA Polymerases from Thermophilic *Geobacillus* species Isolated from Hot Springs in Turkey: Characterization of a DNA Polymerase I from *Geobacillus kaue* Strain NB. *Applied Biochemistry and Biotechnology.* **165**(5):1188-1200.

**Campanella** J.J, Bitincka L, Smalley J. 2003. MatGAT: An application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics*. **4**:29.

**Capson** T.L, Peliska J.A, Kaboord B.F, Frey M.W, Lively C, Dahlberg M, Benkovic S.J. 1992. Kinetic characterisation of the polymerase and exonuclease activities of the gene 43 protein of bacteriophage. *Biochemistry*. **31**(45):10984-94.

**Chen** Y, Cai S, Zhu L, Zhang Z. 2007. Determination of the three-dimensional structure of the Mrf2-DNA complex using paramagnetic spin labelling. *Biochemistry* **46**:4943-4950.

**Clark** J.M. 1988. Novel non-templated nucleotide addition reactions catalyzed by procaryotic and eucaryotic DNA polymerases. *Nucleic Acids Res*. **16**:9677- 9686.

**Clark** D.R., Morant N.R. GeneSys Biotech Ltd. 2009. Enzyme. Patent: WO2009106795.

**Clark** D.R, Wilkinson M, Morant N.R. GeneSys Biotech Ltd. 2009. Cren7 chimeric protein. Patent application: WO2009087394.

**Compton** J. 1991. Nucleic acid sequence-based amplification. *Nature*. **350**(6313):91-92.

**Craw** P. Balachandran W. 2012. Isothermal nucleic acid amplification technologies for point-of-care diagnostics: a critical review. *Lab Chip*. **12**(14):2469-2486.

**Danson** M.J, Hough D.W, Russell R.J, Taylor G.L, Pearl L. 1996. Enzyme thermostability and thermoactivity. *Protein Eng*. **9**(8):629-30.

**Danson** M.J, Daniel R.M, Hough D.W, Lee C.K, Peterson M.E, Cowan D.A. 2008. Protein Adaptation in Extremophiles. [Eds: K.S. Siddiqui & T. Thomas] p1-34. Nova Science Publishers, New York.

**Das** R, Gerstein M. 2000. The stability of thermophilic proteins: a study based on comprehensive genome comparison. *Funct Integr Genomics*. **1**:76-88.

**Davidson** J.F, Fox R, Harris D.D, Lyons-Abbott S, Loeb L.A. 2003. Insertion of the T3 DNA polymerase thioredoxin binding domain enhances the processivity and fidelity of *Taq* DNA polymerase. *Nucleic Acids Res*. **31**(16):4702-9.

**Dean** F.B, Nelson J.R, Giesler T.L, Lasken R.S. 2001. Rapid Amplification of Plasmid and Phage DNA Using Phi29 DNA Polymerase and Multiply-Primed Rolling Circle Amplification. *Genome research*. **11**(6):1095–1099.



**Delarue** M, Poch O, Tordo N, Moras D, Argos P. 1990. An attempt to unify the structure of polymerases. *Protein Engineering*. **3**(6):461-467.

**Derbyshire** V, Freemont P.S, Sanderson M.R, Beese L, Friedman J.M, Joyce C.M, Steitz T.A. 1988. Genetic and crystallographic studies of the 3'-5' exonucleolytic site of DNA polymerase I. *Science*. **240**(4849):199-201.

**Derbyshire** V, Grindley N.D, Joyce C.M. 1991. The 3'-5' exonuclease of DNA polymerase I of *Escherichia coli*: contribution of each amino acid at the active site to the reaction. *EMBO J*. **10**(1):17–24.

**Derbyshire** V, Pinsonneault J.K, Joyce C.M. 1995. Structure-function analysis of 3' to 5' exonuclease of DNA polymerases. *Methods Enzymology*. **262**:363-385.

**Dipippo** J.L, Nesbø C.L, Dahle H, Doolittle W.F, Birkland N.K, Noll K.M. 2009. *Kosmotoga olearia* gen. nov., sp. nov., a thermophilic, anaerobic heterotroph isolated from an oil production fluid. *Int J Syst Evol Microbiol*. **59**(12):2991-3000.

**Doublé** S, Tabor S, Long A.M, Richardson C.C, Ellenberger T. 1998. Crystal structure of a bacteriophage T7 DNA replication complex at 2.2 Å resolution. *Nature*. **391**:251-258.

**Doublé** S, Sawaya M.R, Ellenberger T. 1999. An open closed case for all polymerases. *Structure*. **7**:31-35.

**Fire** A, Xu S.Q. 1995. Rolling replication of short DNA circles. *Proc Natl Acad Sci USA*. **92**(10):4641–4645.

**Friedberg** E.C, Feaver W.J, Gerlach V.L. 2000. The many faces of DNA polymerases: strategies for mutagenesis and for mutational avoidance. *Proc Natl Acad Sci USA*. **97**(11):5681-5683.

**Geojith** G, Dhanasekaran S, Chandran S.P, Kenneth J. 2011. *Efficacy of loop mediated isothermal amplification (LAMP) assay for the laboratory identification of Mycobacterium tuberculosis isolates*. *J Microbiol Methods*. **84**(1):71-73.

**Gill** P. Ghaemi A. 2008. Nucleic acid isothermal amplification technologies: a review. *Nucleosides Nucleotides Nucleic Acids*. **27**(3):224-243.

**Goodman** M.F, Fygenon K.D. 1998. DNA polymerase fidelity: from genetics toward a biochemical understanding. *Genetics*. **148**(4):1475–1482.

**Goodman** M.F, Tang M, Shen X, Frank E.G, O'Donnell M, Woodgate R. 1999. UmuD'<sub>2</sub>C is an error-prone DNA polymerase, *Escherichia coli* DNA pol V. *Proc Natl Acad Sci USA*. **96**:8919–8924.

**Götz** D, Banta A, Beveridge T.J, Rushdi A.I, Simoneit B.R, Reysenbach A.L. 2002. *Persephonella marina* gen. nov., sp. nov. and *Persephonella guaymasensis* sp. nov., two novel thermophilic hydrogen-oxidizing microaerophiles from deep-sea hydrothermal vents. *Int J Syst Evol Microbiol.* **52**: 1349–1359.

**Guex** N, Peitsch M C. 1997. SWISS-MODEL and the Swiss-PdbViewer: An environment for comparative protein modeling. *Electrophoresis* **18**:2714-2723.

**Guex** N, Peitsch M.C, Schwede T. 2009. Automated comparative protein structure modeling with SWISS-MODEL and Swiss-PdbViewer: a historical perspective. *Electrophoresis.* **30**:S162-73.

**Guoliang** F. Oxitec Ltd. 2012. Strand displacement activity of modified polymerases and uses thereof. Patent: US20120115145 A1.

**Hamilton** S.C, Farchaus J.W, Davis M.C. 2001. DNA Polymerases as engines for biotechnology *Biotechniques.* **31**:370-383.

**Harris** C.L, Sanchez-Vargas I.J, Olson K.E, Alphey L, Fu G.. 2013. Polymerase chain displacement reaction. *Biotechniques.* **54**(2):93-97.

**Henneke** G, Flament D, Hübscher U, Querellou J, Raffin J.P. 2005. The hyperthermophilic euryarchaeota *Pyrococcus abyssi* likely requires the two DNA polymerases D and B for DNA replication. *J Mol Biol.* **350**:53-64.

**Hollingsworth** H.C, Nossal N.G. 1991. Bacteriophage T4 encodes an RNase H which removes RNA primers made by the T4 DNA replication system in vitro. *J Bio Chem* **266**:1888-97.

**Holton** T.A, Graham M.W. 1991. A simple and efficient method for direct cloning of PCR products using ddT-tailed vectors. *Nucleic Acids Res.* **19**:1156.

**Horton** R.M, Ho S.N, Hunt H.D, Pullen J.K, Pease L.R. 1989. Site-directed mutagenesis by overlap extension using the polymerase chain reaction. *Gene.* **77**(1):51-9.

**Hough** D.W, Danson M.J. 1999. Extremozymes. *Curr. Opin. Chem. Biol.* **3**:39-46.

**Guo** L, Feng Y, Zhang Z, Yao H, Luo Y, Wang J, Huang L. 2008. Biochemical and structural characterization of Cren7, a novel chromatin protein conserved among Crenarchaea. *Nucleic Acids Res.* **36**(4):1129–1137.

**Huang** S, Sheng P, Zhang H. 2012. Isolation and Identification of Cellulolytic Bacteria from the Gut of *Holotrichia parallela* Larvae. *Int J Mol Sci.* **13**(3):2563-2577.

**Huang** L, Sun F. 2013. *Sulfolobus* chromatin proteins modulate strand displacement by DNA polymerase B1. *Nucl. Acids Res.* **41**(17):8182-8195.

**Ishino** Y, Iwasaki H, Kato I, Shinagawa H. 1994. Amino acid sequence motifs essential to 3'-5' exonuclease activity of *Escherichia coli* DNA polymerase II. *J Biol Chem.* **269**:14655-14660.

**Ito** J, Braithwaite D.K. 1991. Compilation and alignment of DNA polymerase sequences. *Nucleic acids research.* **19**(15):4045-4057.

**Jeanthon** C, L'Haridon S, Cuff V, Banta A, Reysenbach A.L, Prieur D. 2002. *Thermodesulfobacterium hydrogenophilum* sp. nov., a thermophilic, chemolithoautotrophic, sulfate-reducing bacterium isolated from a deep-sea hydrothermal vent at Guaymas Basin, and emendation of the genus *Thermodesulfobacterium*. *Int. J. Syst. Evol. Microbiol.* **52**(3):765–772.

**Joyce** C.M, Steitz T.A. 1987. DNA polymerase I: from crystal structure to function via genetics. *Trends Biochem Sci.* **12**:288–292.

**Joyce** C.M, Steitz T.A. 1994. Function and structure relationships in DNA Polymerases. *Annu. Rev. Biochem.* **63**:777-822.

**Joyce** C.M, Steitz T.A. 1995. Polymerase structure and function: Variations on a theme? *Journal of Bacteriology.* **177**(22) 6321-6329.

**Joyce** C.M, Turner R.M, Grindley N.D. (2003). Interaction of DNA polymerase I (Klenow fragment) with the single-stranded template beyond the site of synthesis. *Biochemistry.* **42**(8):2373-85.

**Kaboev** O.K, Luchkina L.A, Akhmedov A.T, Bekker M.L. 1981. Purification and properties of deoxyribonucleic acid polymerase from *Bacillus stearothermophilus*. *J Bacteriol.* **145**(1):21–26.

**Kanaya** S, Oobatake M, Liu Y. 1996. Thermal Stability of *Escherichia coli* Ribonuclease HI and Its Active Site Mutants in the Presence and Absence of the Mg<sup>2+</sup> Ion - proposal of a novel catalytic role for Glu<sup>48\*</sup>. *J. Biol. Chem.* **271**(51):32729–32736.

**Kaneko** H, Kawana T, Fukushima E, Suzutani T. 2007. Tolerance of loop-mediated isothermal amplification to a culture medium and biological substances. *J Biochem Biophys Methods.* **70**(3):499-501.

**Kiefer** J.R, Mao C, Hansen C.J, Basehore S.L, Hogrefe H.H, Braman J.C, Beese L.S. 1997. Crystal structure of a thermostable *Bacillus* DNA polymerase I large fragment at 2.1 Å resolution. *Structure*. **5**:95-108.

**Kiefer** J.R, Mao C, Braman J.C, Beese L.S. 1998. Visualizing DNA replication in a catalytically active *Bacillus* DNA polymerase crystal. *Nature*. **391**(6664):304-7.

**Klenow** H, Henningsen I. 1970. Selective elimination of the exonuclease activity of the deoxyribonucleic acid polymerase from *E.coli* B by limited proteolysis. *Proc Natl Acad Sci USA*. **65**:168-175.

**Kong** H, Kucera R.B, Jack W.E. 1993. Characterization of a DNA polymerase from the hyperthermophile archaea *Thermococcus litoralis*. Vent DNA polymerase, steady state kinetics, thermal stability, processivity, strand displacement, and exonuclease activities. *J Biol Chem*. **268**(3):1965-75.

**Kornberg** A, Lehman I.R, Bessman M.J, Simms E.S. 1958. Enzymatic Synthesis of Deoxyribonucleic Acid. I. Preparation of Substrates and Partial Purification of an Enzyme from *Escherichia coli*. *J. Biol. Chem*. **233**:163-170

**Kornberg** A, Scott J.F. Bertsch L.L. 1978. ATP utilization by rep protein in the catalytic separation of DNA strands at a replicating fork. *J Biol Chem*. **253**:3298–304

**Kornberg** A. 1980. DNA Replication. W.H. Freeman, San Francisco.

**Kornberg** A. Baker. 1992. DNA Replication. W. H. Freeman, New York.

**Kornberg** A. 1992. Prereplicative complexes of components of DNA polymerase III holoenzyme of *Escherichia coli*. *J Biol Chem*. **267**(15):10318–10322.

**Korolev** S, Nayal M, Barnes W.M, DiCera E, Waksman G. 1995. Crystal structure of the large fragment of *Thermus aquaticus* DNA polymerase I at 2.5 Å resolution: Structural basis for thermostability, *Proc. Natl. Acad. Sci*. **92**:9264-9268.

**Kunkel** T.A, Eckert K.A. 1990. High fidelity DNA synthesis by the *Thermus aquaticus* DNA polymerase. *Nucleic Acids Res*. Jul 11, 1990; **18**(13): 3739–3744.

**Kunkel** T.A. Bebenek K. 2000. DNA replication fidelity. *An rev Biochem*. **69**:497-529.

**Kunkel** T.A. McCulloch S.D. 2008. The fidelity of DNA synthesis by eukaryotic replicative and translesion synthesis polymerases. *Cell Res*. **18**(1):148-61.

KRX manual, Single Step (KRX) Competent Cells Technical Bulletin, Promega, USA, 2010.

**Kumar** L, Awasthi G, Singh B. 2011. Extremophiles: A Novel Source of Industrially Important Enzymes. *Biotechnology*. **10**:121-135.

**Lane** D.J, Pace B, Olsen G.J, Stahl D.A, Sogin M.L, Pace N.R. 1985. Rapid determination of 16S ribosomal RNA sequences for phylogenetic analyses. *Proc Natl Acad Sci USA*. **82**(20):6955-9.

**Lane** D.J, Weisburg W.G, Barns S.M, Pelletier D.A. 1991. 16S ribosomal DNA amplification for phylogenetic study. *J Bacteriol*. **173**(2):697-703.

**Larkin** M.A, Blackshields G, Brown N.P, Chenna R, McGettigan P.A, McWilliam H, Valentin F, Wallace I.M, Wilm A, Lopez R, Thompson J.D, Gibson T.J and Higgins D.G. 2007. ClustalW and ClustalX version 2 (2007). *Bioinformatics* **23**(21): 2947-2948.

**Lee** Y.J, Wagner I.D, Brice M.E, Kevbrin V.V, Mills G.L, Romanek C.S, Wiegel J. 2005. *Thermosediminibacter oceani* gen. nov., sp. nov. and *Thermosediminibacter litoriperuensis* sp. nov., new anaerobic thermophilic bacteria isolated from Peru Margin. *Extremophiles*. **9**:375-383.

**Lehmann** I.R, Nussbaum A.L. 1964. The deoxyribonucleases of *E.coli* V – on the specificity of exonuclease I. *J. Biol. Chem*. **239**(8):2628-2636.

**Loeb** L.A, Patel P.H. 2000. DNA polymerase active site is highly mutable evolutionary consequences. *Proc Natl Acad Sci USA*. **97**(10):5095–5100.

**Lowry** O.H, Rosebrough N.J, Farr A.L, Randall R.J. 1951. Protein measurement with the Folin phenol reagent. *J. Biol. Chem*. **193**(1):265–75.

**Li** Y, Korolev S, Waksman G. 1998. Crystal structures of open and closed forms of binary and ternary complexes of the large fragment of *Thermus aquaticus* DNA polymerase I: structural basis for nucleotide incorporation. *EMBO J*. **17**:7514-7525.

**Li** Y, Motré A, Kong H. 2008. Enhancing helicase-dependent amplification by fusing the helicase with the DNA polymerase. *Gene*. **420**:17–22.

**Licata** V, Karantzeni I, Ruiz C, Liu C, 2003. Comparative thermal denaturation of *Thermus aquaticus* and *Escherichia coli* type 1 DNA polymerases. *Biochem J*. **374**(3):785–792.

**Lin** T.C, Karam G, Konigsberg W.H.. 1994. Isolation, characterisation, and kinetic properties of truncated forms of T4 DNA polymerase that exhibit 3'-5' exonuclease activity. *J. Biol. Chem*. **269**:19286-19294.

**Liu D**, Daubendiek S.L, Zillman M.A, Ryan K, Kool E.T.. 1996. Rolling circle DNA synthesis: Small circular oligonucleotides as efficient templates for DNA polymerases. *J Am Chem Soc.* 118:1587–1594.

**Longo M.C**, Berninger M.S, Hartley J.L. 1990. Use of uracil DNA glycosylase to control carry-over contamination in polymerase chain reactions. *Gene.* **93**:125–128

**Lundquist R.C**, Olivera B.M. 1982. Transient generation of displaced single-stranded DNA during nick translation. *Cell.* **31**:53-60.

**Maga G**, Villani G, Tillement V, Stucki M, Locatelli G.A, Frouin I, Spadari S, Hübscher U. 2001. Okazaki fragment processing: modulation of the strand displacement activity of DNA polymerase delta by the concerted action of replication protein A, proliferating cell nuclear antigen, and a flap endonuclease-1. *Proc Natl Acad Sci USA.* **98**:14298-14303.

**Maidak B.L**, Olsen G.J, Larsen N, Overbeek R, McCaughey M.J, Woese C.R. 1997. The RDP (Ribosomal Database Project). *Nucleic Acids Research.* **25**(1):109-110.

**Marceau A.H.** 2012. Functions of single-strand DNA-binding proteins in DNA replication, recombination, and repair. *Methods Mol Biol.* **922**:1-21.

**Metzler R**, Ambjörnsson T, Hanke A, Fogedby H.C. 2009. Single DNA denaturation and bubble dynamics. *J Phys: Condens Matter.* **21**:034111 .

**Mitsunaga S**, Shimizu S, Okudaira Y, Oka A, Tanaka M, Kimura M, Kulski J.K, Inoue I, Inoko H.. 2013. Improved loop-mediated isothermal amplification for HLA-DRB1 genotyping using RecA and a restriction enzyme for enhanced amplification specificity. *Immunogenetics.* **65**(6):405-15.

**Mori Y**, Notomi T. 2009. Loop-mediated isothermal amplification (LAMP): a rapid, accurate, and cost-effective diagnostic method for infectious diseases. *J Infect Chemother.* **15**(2):62-9.

**Moussard H**, L'Haridon S, Tindall B.J, Banta A, Schumann P, Stackebrandt E, Reysenbach A.L, Jeanthon C.. 2004. *Thermodesulfatator indicus* gen. nov., sp. nov., a novel thermophilic chemolithoautotrophic sulfate-reducing bacterium isolated from the Central Indian Ridge. *Int J Syst Evol Microbiol.* **54**:227–233.

**Muller P.Y**, Studer E, Miserez A.R. 2001. Molecular Biocomputing Suite: a word processor add-in for the analysis and manipulation of nucleic acid and protein sequence data. *Biotechniques.* **31**(6):1306-1313.

**Mullis** K.B, Faloona F.A. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods Enzymol* **155**:335-50.

**Myers** T.W, Gelfand D.H. 1991. Reverse transcription and DNA amplification by a *Thermus thermophilus* DNA polymerase. *Biochemistry*. **30**(31):7661-6.

**Mori** Y, Nagamine K, Tomita N, Notomi T. 2001. Detection of Loop-Mediated Isothermal Amplification Reaction by Turbidity Derived from Magnesium Pyrophosphate Formation. *Biochemical and Biophysical Research Communications*. 289(1):150–154.

**Nagamine** K, Hase T, Notomi T. 2002. Accelerated reaction by loop-mediated isothermal amplification using loop primers. *Molecular and Cellular Probes*. **16**:223-229.

**Nazina** T.N, Tourova T.P, Poltarau A.B, Novikova E.V, Grigoryan A.A, Ivanova A.E, Lysenko A.M, Petrunyaka V.V, Osipov G.A, Belyaev S.S, Ivanov M.V. 2001. Taxonomic study of aerobic thermophilic *bacilli*. *Int J Syst Evol Microbiol*. **51**(2):433-46.

**Niemz** A, Ferguson T.M, Boyle D.S.. 2011. Point-of-care nucleic acid testing for infectious disease. *Trends Biotechnol*. **29**(5):240-250.

**Nixon** A.E, Ostermeier M, Benkovic S.J. 1998. Hybrid enzymes: manipulating enzyme design. *Trends Biotechnology*. **16**(6):258-64.

**Notomi** T, Okayama H, Masubuchi H, Yonekawa T, Watanabe K, Amino N, Hase T. 2000. Loop-mediated isothermal amplification of DNA. *Nucleic Acids Res*. **28**(12):E63.  
Novagen pET<sup>®</sup> System manual, The gold standard for protein expression. 10<sup>th</sup> edition, 2003.

**Novy** R, Drott D, Yaeger K, Mierenhof R. 2001. Overcoming the codon bias of *E. coli* for enhanced protein expression. *inNovations*. **12**:1-3.

**Ollis** D.L, Brick P, Hamlin R, Xuong N.G, Steitz T.A. 1985. Structure of large fragment of *Escherichia coli* DNA polymerase I complexed with dTMP. *Nature*. **6**(313):762-6.

**Oren** A. Diversity of halophilic microorganisms: environments, phylogeny, physiology, and applications. *J Indust Micro Biotch*. 28:56-63.

**Patel** P.H, Suzuki M, Adman E, Shinkai A, Loeb LA. 2001. Prokaryotic DNA Polymerase I: Evolution, Structure, and “Base Flipping” mechanisms for Nucleotide selection. *J. Mol. Biol* **308**:823-837.

**Patel** P.H, Loeb L.A. 2001. Getting a grip on how DNA polymerases function. *Nature structural biology*. **8**(8):656-659.

**Pavlov** A.R, Belova G.I, Kozyavkin S.A, Slesarev A.I. 2002. Helix–hairpin–helix motifs confer salt resistance and processivity on chimeric DNA polymerases. *Proc Natl Acad Sci*. **99**(21):13510-13515.

**Pavlov** A.R, Pavlova N.V, Kozyavkin S.A, Slesarev A.I. 2012. Cooperation between catalytic and DNA binding domains enhances thermostability and supports DNA synthesis at higher temperatures by thermostable DNA polymerases. *Biochemistry*. **51**(10):2032-43.

**Perler** F.B, Southworth M.W, Kong H, Kucera R.B, Ware J, Jannasch H.W. 1996. Cloning of thermostable DNA polymerases from hyperthermophilic marine Archaea with emphasis on *Thermococcus* sp. 9 degrees N-7 and mutations affecting 3'-5' exonuclease activity. *Proc Natl Acad Sci USA*. **93**(11):5281–5285.

**Perler** F.B, Kumar S, Kong H. 1996. Thermostable DNA polymerases. *Adv Prot Chem*. **48**:377-435.

**Pikuta** E, Lysenko A, Chuvilskaya N, Mendrock U, Hippe H, Suzina N, Nikitin D, Osipov G, Laurinavichius K. 2000. *Anoxybacillus pushchinensis* gen. nov., sp. nov., a novel anaerobic alkaliphilic, moderately thermophilic bacterium from manure, and description of *Anoxybacillus flavithermus* comb. nov. *Int. J. Syst. Evol. Microbiol*. **50**:2109-2117.

**Rogozin** I.B, Makarova K.S, Pavlov Y.I, Koonin E.V.. 2008. A highly conserved family of inactivated archaeal B family DNA polymerases. *Biol Direct*. **3**(32).

**Rose** T.M, Schultz E.R, Henikoff J.G, Pietrovski S, McCallum C.M, Henikoff S. 1998. Consensus-degenerate hybrid oligonucleotide primers for amplification of distantly related sequences. *Nucleic Acids Research*. **26**:1628–1635.

**Rose** T.M, Henikoff J.G, Henikoff S. 2003. CODEHOP (COnsensus-DEgenerate Hybrid Oligonucleotide Primer) PCR primer design. *Nucleic. Acids Research*. **31**(13):3763-3766.

**Rothwell** P.J, Waksman G. 2005. Structure and Mechanism of DNA Polymerases. *Advances in Protein Chemistry*. **71**:401-440.

**Roux** K.H. 2009. Optimization and troubleshooting in PCR. *Cold Spring Harb Protoc*. (4).

**Saiki** R.K, Gelfand D.H, Stoffel S, Scharf S.J, Higuchi R, Horn G.T, Mullis K.B, Erlich H.A. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. *Science*. **239**(4839):487–491.



**Saiki** R.K, Higuchi R, Krummel B. 1988. A general method of *in vitro* preparation and specific mutagenesis of DNA fragments: study of protein and DNA interactions. *Nucleic Acids Research*. **16**(15):7351–7367.

**Sako** Y, Nakagawa S, Takai K, Horikoshi K. 2003. *Marinithermus hydrothermalis* gen. nov., sp. nov., a strictly aerobic, thermophilic bacterium from a deep-sea hydrothermal vent chimney. *Int J Syst Evol Microbiol*. **53**:59-65.

**Salas** M, Vega M, Lázaro J, Mencía M, Blanco L. 2010. Improvement of Phi29 DNA polymerase amplification performance by fusion of DNA binding motifs. *Proc Natl Acad Sci USA*. **107**(38):16506–16511.

**Sandalli** C, Singh K, Modak M.J, Ketkar A, Canakci S, Demir I, Belduz A.O. 2009. A new DNA polymerase I from *Geobacillus caldoxylosilyticus* TK4: cloning, characterization, and mutational analysis of two aromatic residues. *Applied Microbiology and Biotechnology*. **84**(1):105-117.

**Savino** C, Federici L, Johnson K.A, Vallone B, Nastopoulos V, Rossi M, Pisani F.M, Tsernoglou D. 2004. Insights into DNA replication: the crystal structure of DNA polymerase B1 from the archaeon *Sulfolobus solfataricus*. *Structure*. **12**:2001–2008.

**Sellmann** E, Schröder K.L, Knoblich I.M, Westermann P. 1992. Purification and characterization of DNA polymerase from *Bacillus* species. *J Bacteriol*. **174**:4350-4355.

**Singh** K, Srivastava A, Patel S.S, Modak M.J. 2007. Participation of the Finger subdomain of *Escherichia coli* DNA polymerase I in the strand displacement synthesis of DNA. *J. Biol. Chem*. **282**:10594-10604.

**Slobodkin** A, Reysenbach A.L, Strutz N, Dreier M, Wiegel J. 1997. *Thermoterrabacterium ferrireducens* gen. nov., sp. nov. a thermophilic anaerobic, dissimilatory Fe(III)-reducing bacterium from a continental hot spring. *Int J Syst Bacteriol*. **47**:541–547.

**Soengas** M.S, Esteban J.A, Lázaro J.M, Bernad A, Blasco M.A, Salas M, Blanco L. 1992. Site-directed mutagenesis at the Exo III motif of phi 29 DNA polymerase; overlapping structural domains for the 3'-5' exonuclease and strand-displacement activities. *EMBO J*. **11**:4227-37.

**Spangler** R, Goddard N.L, Thaler D.S. 2009. Optimizing Taq Polymerase Concentration for Improved Signal-to-Noise in the Broad Range Detection of Low Abundance Bacteria. *PLOS one*. **4**(9):e7010.

**Steitz** T.A, Kim Y, Eom S.H, Wang J, Lee D.S, Suh S.W. 1995. Crystal structure of *Thermus aquaticus* DNA polymerase. *Nature*. **376**:612-616.

**Steitz** T.A, Eom S.H, Wang J. 1995. Structure of *Taq* polymerase with DNA at the polymerase active site. *Nature*. **382**:278-281.

**Steitz** T.A. 1999. DNA Polymerases: Structural Diversity and Common Mechanisms, *J. Biol. Chem.* **274**(18):17395–17398.

**Steitz** T.A, Zhou B.L, Pata J.D. 2001. Crystal structure of a *DinB* lesion bypass DNA polymerase catalytic fragment reveals a classic polymerase catalytic domain. *Molecular Cell*. **2**:427-37.

**Steitz** T.A, Yin Y.W. 2004 Accuracy, lesion bypass, strand displacement and translocation by DNA polymerases. *Phil. Trans. R. Soc. Lond. Biol Sci.* **359**:17-23.

**Stemmer** W.P. 1994. Rapid evolution of a protein *in vitro* by DNA shuffling. *Nature*. **370**(6488):389-91.

**Stenesh** J, Roe. 1972. DNA Polymerase from Mesophilic and Thermophilic Bacteria I - Purification and Properties of the Enzyme. *Biochem. Biophys. Acta*. **272**:156-166.

**Stetter** K.O, Huber R, Langworthy T.A, König H, Sleytr U, Thomm M, Woese C.R. 1986. *Thermotoga maritima* sp. nov. represents a new genus of unique extremely thermophilic eubacteria growing up to 90°C. *Archives of Microbiology*. **144**(4):324-333.

**Stetter** K.O, Huber R, Kristjansson J.K. 1987. *Pyrobaculum* gen. nov., a new genus of neutrophilic, rod-shaped archaeobacteria from continental solfataras growing optimally at 100°C. *Arch. Microbiol.* **149**(2):95-101.

**Suzuki** R, Ihira M, Enomoto Y, Yano H, Maruyama F, Emi N, Asano Y, Yoshikawa T. 2010. Heat denaturation increases the sensitivity of the cytomegalovirus loop-mediated isothermal amplification method. *Microbiology and Immunology*. **54**(8):466–470.

**Sweasy** J.B, Chen M, Loeb L.A. 1992. Mammalian DNA polymerase beta can substitute for DNA polymerase I during DNA replication in *Escherichia coli*. *K. Mol Chem* **267**:1407-1410.

**Tabor** S, and Richardson C.C. 1995. A single residue in DNA polymerase of *Escherichia coli* DNA polymerase I family is critical for distinguishing between deoxy- and dideoxy- ribonucleotides. *Proc Natl Acad Sci USA*. **92**:6339-6343.

**Takami** H, Inoue A, Fuji F, Horikoshi K. 1997. Microbial flora in the deepest sea mud of the Mariana Trench. *FEMS Microbiol Lett*. **152**:279-285.

**Tang L**, Gao H, Zhu X, Wang X, Zhou M, Jiang R. 2012. Construction of "small-intelligent" focused mutagenesis libraries using well-designed combinatorial degenerate primers. *Biotechniques*. **52**(3):149-58.

**Tanner N**, Evans T.C. NEB. 2013. Reducing Template Independent Primer Extension and Threshold Time for Loop Mediated Isothermal Amplification. Patent: US 20130122551.

**Thompson J.D**, Higgins D.G, Gibson T.J. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research*. **22**:4673-4680.

**Thompson M.J**, Eisenberg D. 1999. Transproteomic evidence of a loop-deletion mechanism for enhancing protein thermostability. *J Mol Biol*. **292**(4):946.

**Triglia T**, Peterson M.G, Kemp D.J. 1989. A procedure for in vitro amplification of DNA segments that lie outside the boundaries of known sequences. *Nucleic Acids Research*. **16**(16):8186.

**Uemori T**, Ishino Y, Fujita K, Asada K, Kato I. 1993. Cloning of the DNA polymerase gene of *Bacillus caldotenax* and characterization of the gene product. *J Biochem*. 113:401-10.

**Uemori T**, Sato Y, Kato I, Doi H, Ishino Y. 1997. A novel DNA polymerase in the hyperthermophilic archaeon, *Pyrococcus furiosus*: Gene cloning, expression and characterisation. *Genes to Cell*. **2**:499-512.

**Viguera E**, Canceill D, Ehrlic S. 2001. *In vitro* replication slippage by DNA polymerases from thermophilic organisms. *J Mol Biol*. **312**:323-333.

**Vincent M**, Xu Y, Kong H. 2004. Helicase-dependent isothermal DNA amplification. *EMBO Rep*. **5**(8):795–800.

**Villbrandt B**, Sobek H, Frey B, Schomburg D. 2000. Domain exchange: chimeras of *Thermus aquaticus* DNA polymerase, *Escherichia coli* DNA polymerase I and *Thermotoga neapolitana* DNA polymerase. *Protein Eng*. **9**:645-54.

**Vlasov P.K**, Povolotskaya I.S, Kondrashov F.A, Ledda A. 2012. Stop codons in bacteria are not selectively equivalent. *Biology Direct*. **7**:30.

**Rothwell P.J**, Mitaksov V, Waksman G. 2005. Motions of the Fingers Subdomain of KlenTaq1 Are Fast and Not Rate Limiting: Implications for the Molecular Basis of Fidelity in DNA Polymerases. *Molecular cell*. **19**(3):3345–355.

- Walker** G, Little M, Nadeau J, Shank D. 1992. Isothermal *in vitro* amplification of DNA by a restriction enzyme/DNA polymerase system. *Proc Natl Acad Sci USA*. **89**:392–396.
- Watson** J.D, **Crick** F.H.C. 1953. Structure for deoxyribose nucleic acid. *Nature* **171**:737-738.
- Wang** Y, Prosen D.E, Mei L, Sullivan J.C, Finney M, Vander Horn P.B. 2004. A novel strategy to engineer DNA polymerases for enhanced processivity and improved performance *in vitro*. *Nucleic Acids Res*. **32**(3):1197–1207.
- Wartell** R.M, Benight A.S. 1985. Thermal denaturation of DNA molecules: A comparison of theory with experiment. *Phys Rep*. **126**:67–107 (1985).
- Waugh** D.S, Sun P, Tropea J.E. 2011. Enhancing the solubility of recombinant proteins in *Escherichia coli* by using hexahistidine-tagged maltose-binding protein as a fusion partner. *Methods Mol. Biol*. **705**:259-274.
- Wittwer** C.T, Pryor R.J. 2006. Real-time polymerase chain reaction and melting curve analysis. *Methods Mol Biol*. **336**:19-32.
- Woese** C.R. 1987. Bacterial evolution. *Microbiol Rev*. **51**(2):221-71.
- Woodgate** R, Ling H, Boudsocq F, Yang W. 2001. Crystal structure of a Y-family DNA polymerase in action: a mechanism for error-prone and lesion-bypass replication. *Cell*. **107**(1):91-102.
- Woodgate** R, Boudsocq F, Iwai S, Hanaoka F. 2001. *Sulfolobus solfataricus* P2 DNA polymerase IV (Dpo4): an archaeal DinB-like DNA polymerase with lesion-bypass properties akin to eukaryotic polη. *Nucleic Acids Res*. **29**:4607-4616.
- Xie** P. 2012. Modeling translocation dynamics of strand displacement DNA synthesis by DNA polymerase I. *J Mol Model*. **18**:1951-1960.
- Xu** G, Hu L, Zhong H, Wang H, Yusa S, Weiss T, Romineuk P, Pickerill S, You Q. 2012. Cross Priming Amplification: Mechanism and Optimization for Isothermal DNA Amplification. *Sci Rep*. **2**:246.
- Yang** S.W, Astatke M, Potter J, Chatterjee D.K. 2002. Mutant *Thermotoga neapolitana* DNA polymerase I: altered catalytic properties for non-templated nucleotide addition and incorporation of correct nucleotides. *Nucleic Acids Res*. **30**(19):4314–4320.

**Yuan** Y.C, Whitson R.H, Liu Q, Itakura K, Chen Y. 1998. A novel DNA-binding motif shares structural homology to DNA replication and repair nucleases and polymerases. *Nature structural biology*. **5**:959-964.

**Zhu** L, Hu J, Lin D, Whitson R, Itakura K, Chen Y. 2001. Dynamics of the Mrf-2 DNA-Binding Domain Free and in Complex with DNA. *Biochemistry*. **40**(31):9142–9150.

## V. APPENDIX

### i. Primer sequences

#### i. 16S rRNA Primers: (Lane *et al.*, 1991)

27f	5'-AGAGTTTGATCMTGGCTCAG-3'
1492r	5'-TACGGYTACCTTGTTACGACTT-3'

#### ii. CODEHOP primers: (Bergquist *et al.*, 2004)

PolATF	5'-CATTTTTGCTGCCGATTAYWSNCARATHGA-3'
PolATR	5'-TTGGCGCTTCAAAAATAARYTCRTCRTG-3'
PolGCF1	5'-ATGGCCGCCGACTACTCNCARATHGA-3'
PolGCF2	5'-ATGGCCGCCGACTACAGYCARATHGA-3'
PolGCR	5'-TGGGCACCTCGAAGATCARYTCRTCRTG-3'

#### iii. Vector specific primers:

M13_Reverse	5'-AGCGGATAACAATTTACACAGGA-3'
M13_Universal	5'-TGTAACACGACGGCCAGT-3'
T7_Promoter	5'-AAATACGCTTCACTATAGGG-3'
T7_Terminator	5'-GCTAGTTATTGCTCAGCGG-3'

#### iv. Gene-walking primers:

Gme_GW_U1	5'-GTCGTCATCGGCGATATGGGCGAG-3'
Gme_GW_UN1	5'-ATCCAAATCGCGTTGGAACGC-3'
Gme_GW_LN1	5'-gATTGATTTAGCGGCACGGCTG-3'
Gme_GW_L1	5'-AGCTTCAGGCTCGTCTTTTGCTG-3'
Bst_777_U	5'-CGTCTACGAAGGACAAGACCGCGA-3'
Gme_GWL2	5'-TCGCGGTCTTGTCTTTCGTAGACG-3'

Tin_GW2_UN	5'-CATGCGTGGGAAAATAGTA-3'
Tat_GW_NterU	5'-CACTCATGGCCTTGATGA-3'
Tat_GW_NterL	5'-GTCGTATTTTATGTTGTGGCC-3'
Tat_GW_Nter_Un	5'-ACAAAAGAACTTGCCAAAGGAGA-3'
Tat_GW_Nter_Ln	5'-CTTGTTGGTCCTCGATGAGG-3'
Tin_(pos-154)U	5'-CGCGCATTACGTGGTCAA -3'

Gcx_487_U:	5'-CGCTGAGCAGGAAGGGTTTGAG-3'
Gcx_Cter_L:	5'-ACATGGTATGACGCGAAATGA-3'
Cfe_LF_U	5'-ATGAAAGAAATTGAACCAAAAAATAAGAAAGA-3'
Cfe_C-ter_L	5'-TTACTTCGCTTCATACCAG-3'
Thy_GW_U	5'-TATTCAGGGCTCAGCCTCTG-3'
Thy_GW_Un	5'-AAAGTGTGCTATGGTGGCTC-3'
Thy_GW_L	5'-ACACCAAAAACCTCACAAGC-3'
Thy_GW_Ln	5'-TCACAAGCAGTAAAAAGTATGAATATC-3'
Thy_GW2_U	5'-GAATTAACAGTTGCCTCTGATG-3'
Thy_GW2_Un	5'-CTCTTTTCCTTAACCAATCCTG-3'
Thy_GW2_L	5'-TCGACAGAAAGTTTATAAGCCT-3'
Thy_GW2_Ln	5'-CATGAAGAATAAATTCAGACTTAGC-3'

**v. Cloning primers** : (restriction enzymes are boxed)

GmePol_LF_F	5'-GCGCCG <b>CATATG</b> GAAGGGGAGAAACCGCTTGAGATG-3'
GmePol_R	5'-TTTCCG <b>GGATCC</b> TTATTTGGCATCATACCATGTTG-3'
GkaPol_LF_F	5'-GAATTC <b>CATATG</b> TCAGAAGAGGAAAAACCGCTTGC-3'
GkaPol_R	5'-AAGCTT <b>GGATCC</b> TTATTTTCGCGTCATACCATGTTCG-3'
GcxPol_LF_F	5'-GAATTC <b>CATATG</b> GAAGAGGAAAAACCGCTGGC-3'
GcxPol_R	5'-AAGCTT <b>GGATCC</b> TTATTTTCGCGTCATACCATGTTCG-3'
BfsPol_LF_F	5'-GCGCCG <b>CATATG</b> AAAGAAGAAGAACAGTTGGA-3'
BfsPol_R	5'-AAGCTT <b>GGATCC</b> TTATTTTGCATCGTACCATGAAG-3'
AflPol_LF_F	5'-GAATTC <b>CCATGG</b> AACAAGAACAACATCATTAACCGAT -3'
AflPol_R	5'-AAGCTT <b>GGATCC</b> TTACTTCGCATCATACCACGT-3'
KolPol_LF_F	5'-GAATTC <b>CATATG</b> GCTAACAGGCTTTTTCT-3'
KolPol_R	5'-GAATTC <b>GTCGAC</b> TTATTTACCCAGTAACTGT-3'
CfePol_LF_F	5'-GAATTC <b>CATATG</b> AAAGAAATTGAACCAAAAAATAAG-3'
CfePol_R	5'-AAGCTT <b>GTCGAC</b> TTACTTCGCTTCATACCAG-3'
ThyPol_LF_F	5'-GAATTC <b>CATATG</b> GTTTTTAACAGAATTAAAAATTCCTTCTG-3'
ThyPol_R	5'-AAGCTT <b>GTCGAC</b> TTATTTCTGCCAGTTTTTACCAATG-3'

TinPol_LF_F	5'-CTTGAA <b>CATATG</b> GGCCTCTTAAAAGAACTTCCAGCTAC-3'
TinPol_R	5'-AGCCCT <b>GTCGAC</b> TTATGCCTTTGCCTCTGC-3'
TatPol_LF_F	5'-GAATTC <b>CATATG</b> CTTCTCAAAGAACTCCCCGCCA-3'
TatPol_R	5'-GAATTCGTCGACTTATGCCTTTGCCTCTGCCCAATTTTCCCTAAAGC-3'
TocPol_LF_F	5'-GAATTC <b>CATATG</b> CTCATAAAATAAACTGCCCCGCCCGCAAG3'
TocPol_R	5'-GAATTCGGATCCTTAGCTTATTTCTTCCCAGGTATAGC-3'
MhyPol_LF_F	5'-GAATTC <b>CATATG</b> CTGCGCGAGCTGGGCCTG-3'
MhyPol_R	5'-GAATTC <b>GTCGAC</b> TACTTCGCCTCCAGCCAGTTCTC-3'

#### vi. Small-intelligent primers: (Degenerate codons are identified in bold)

Gme_S717X_NDT_U	5'-acttcggtatcgtttacggaatt <b>NDT</b> gattacggattggcgcaaaactt-3'
Gme_S717X_NDT_L	5'-aagttttgcgccaatccgtaatacahnaattccgtaaaccgataccgaagt-3'
Gme_S717X_VMA_U	5'-acttcggtatcgtttacggaatt <b>VMA</b> gattacggattggcgcaaaactt-3'
Gme_S717X_VMA_L	5'-aagttttgcgccaatccgtaatctkbaattccgtaaaccgataccgaagt-3'
Gme_S717X_ATG_U	5'-acttcggtatcgtttacggaatt <b>ATG</b> gattacggattggcgcaaaactt-3'
Gme_S717X_ATG_L	5'-aagttttgcgccaatccgtaatccataattccgtaaaccgataccgaagt-3'
Gme_S717X_TGG_U	5'-acttcggtatcgtttacggaatt <b>TGG</b> gattacggattggcgcaaaactt-3'
Gme_S717X_TGG_L	5'-aagttttgcgccaatccgtaatcccaaatccgtaaaccgataccgaagt-3'
Gme_Y719X_NDT_U	5'-gtatcgtttacggaattagcgat <b>NDTg</b> gattggcgcaaaacttgaacat-3'
Gme_Y719X_NDT_L	5'-atgttcaagttttgcgccaatccahnatcgctaattccgtaaaccgatac-3'
Gme_Y719X_VMA_U	5'-gtatcgtttacggaattagcgat <b>VMA</b> ggattggcgcaaaacttgaacat-3'
Gme_Y719X_VMA_L	5'-atgttcaagttttgcgccaatccctkbatcgctaattccgtaaaccgatac-3'
Gme_Y719X_ATG_U	5'-gtatcgtttacggaattagcgat <b>ATG</b> ggattggcgcaaaacttgaacat-3'
Gme_Y719X_ATG_L	5'-atgttcaagttttgcgccaatcccatatcgctaattccgtaaaccgatac-3'
Gme_Y719X_TGG_U	5'-gtatcgtttacggaattagcgat <b>TGG</b> ggattggcgcaaaacttgaacat-3'
Gme_Y719X_TGG_L	5'-atgttcaagttttgcgccaatccccaatcgctaattccgtaaaccgatac-3'



Gme_R789X_NDT_U	5'-tcaacgtccgcagttttgcagag <b>NDT</b> acggccatgaacacgccaattca-3'
Gme_R789X_NDT_L	5'-tgaattggcgtgttcatggccgtahnctctgcaaaactgcgagcgttga-3'
Gme_R789X_VMA_U	5'-tcaacgtccgcagttttgcagag <b>VMA</b> acggccatgaacacgccaattca-3'
Gme_Y719X_VMA_L	5'-tgaattggcgtgttcatggccgttkbctctgcaaaactgcgagcgttga-3'
Gme_R789X_ATG_U	5'-tcaacgtccgcagttttgcagag <b>ATG</b> acggccatgaacacgccaattca-3'
Gme_R789X_ATG_L	5'-tgaattggcgtgttcatggccgtcatctctgcaaaactgcgagcgttga-3'
Gme_R789X_TGG_U	5'-tcaacgtccgcagttttgcagag <b>TGG</b> acggccatgaacacgccaattca-3'
Gme_R789X_TGG_L	5'-tgaattggcgtgttcatggccgtccactctgcaaaactgcgagcgttga-3'
Tin_S743X_NDT_U	5'-actttggcattgtttacggcatg <b>NDT</b> ccttacggtctggcgaaagaact-3'
Tin_S743X_NDT_L	5'-agttctttcgccagaccgtaaggahncatgccgtaaacaatgccaaagt-3'
Tin_S743X_VMA_U	5'-actttggcattgtttacggcatg <b>VMA</b> ccttacggtctggcgaaagaact-3'
Tin_S743X_VMA_L	5'-agttctttcgccagaccgtaaggtkbcatgccgtaaacaatgccaaagt-3'
Tin_S743X_ATG_U	5'-actttggcattgtttacggcatg <b>ATG</b> ccttacggtctggcgaaagaact-3'
Tin_S743X_ATG_L	5'-agttctttcgccagaccgtaaggcatcatgccgtaaacaatgccaaagt-3'
Tin_S743X_TGG_U	5'-actttggcattgtttacggcatg <b>TGG</b> ccttacggtctggcgaaagaact-3'
Tin_S743X_TGG_L	5'-agttctttcgccagaccgtaaggccacatgccgtaaacaatgccaaagt-3'
Tin_Y745X_NDT_U	5'-gcattgtttacggcatgagtcct <b>NDT</b> gggtctggcgaaagaactcaaaattg3'
Tin_Y745X_NDT_L	5'-caattttgagttctttcgccagaccahnaggactcatgccgtaaacaatgc3'
Tin_Y745X_VMA_U	5'-gcattgtttacggcatgagtcct <b>VMA</b> gggtctggcgaaagaactcaaaattg3'
Tin_Y745X_VMA_L	5'-caattttgagttctttcgccagacctkbaggactcatgccgtaaacaatgc3'
Tin_Y745X_ATG_U	5'-gcattgtttacggcatgagtcct <b>ATG</b> gggtctggcgaaagaactcaaaattg3'
Tin_Y745X_ATG_L	5'-caattttgagttctttcgccagaccataggactcatgccgtaaacaatgc3'
Tin_Y745X_TGG_U	5'-gcattgtttacggcatgagtcct <b>TGG</b> gggtctggcgaaagaactcaaaattg3'
Tin_Y745X_TGG_L	5'-caattttgagttctttcgccagaccccaaggactcatgccgtaaacaatgc3'
Tin_R815X_NDT_U	5'-gtacggcgcgagagtttgccgag <b>NDT</b> acggctataaaactcctattca-3'
Tin_R815X_NDT_L	5'-tgaataggagtgtttatagccgtahnctcggc aaactcgcgcgccgtac-3'
Tin_R815X_VMA_U	5'-gtacggcgcgagagtttgccgag <b>VMA</b> acggctataaaactcctattca-3'
Tin_R815X_VMA_L	5'-tgaataggagtgtttatagccgttkbctcggc aaactcgcgcgccgtac-3'
Tin_R815X_ATG_U	5'-gtacggcgcgagagtttgccgag <b>ATG</b> acggctataaaactcctattca-3'
Tin_R815X_ATG_L	5'-tgaataggagtgtttatagccgtcatctcggc aaactcgcgcgccgtac-3'
Tin_R815X_TGG_U	5'-gtacggcgcgagagtttgccgag <b>TGG</b> acggctataaaactcctattca-3'
Tin_R815X_TGG_L	5'-tgaataggagtgtttatagccgtccactcggc aaactcgcgcgccgtac-3'

## vii. Characterisation primers:

-47_M13_L	5'-CGCCAGGGTTTTCCCAGTCACGACGTTGTAAAAACGACGGCC-3'
-47_extension	5'-GCCGTTGCTACCCCTCGTTCGATGCTGTC <sup>s</sup> T-3' (C <sup>s</sup> = Phosphorothioate modification)
5kb_blocking	5'-GCCGTTGCTACCCCTCGTTCGATGCTGTCTPHO-3' (PHO = 3' phosphate modification)
Ape_FIP (F1c+F2)	5'-CGTACAGCCTGAGATCGACAAGGTAGTAATGCTAGAGCATCTGG-3'
Ape_BIP (B1c+B2)	5'-GTGTAGACGAAGAGGCGGTGCTCAACACTACAGCCACG-3'
Ape_F3	5'-GGCAATAGTACATGAAGGCT-3'
Ape_B3	5'-CATAAGATTGTCCACGGCTAT-3'
Ape_LoopF	5'-AGCGTGACCTACAGCAAC-3'
Ape_LoopB	5'-AAGGATGAGACGAGAAGTGTG-3'
MS2_FIP (F1c+F2)	5'-TCGCAAGCGAACCATCTACGTCCGTTGGTATAGACCTGAA-3'
MS2_BIP (B1c+B2)	5'-CGATAGACTTATCGTCTGCATCCGAGATATGAATATAGCTCTGGTGG-3'
MS2_F3	5'-TGTAAGGAGCCTGATATGAATATG-3'
MS2_B3	5'-TAGTGTGAGCGGATACGAT-3'
MS2_LoopF	5'-GCCAGACGCTGGTTGAT-3'
MS2_LoopB	5'-GATCGCCTGGTGTGGAG-3'
Lambda_U	5'-GATGAGTTCGTGTCCGTACAACCTGG-3'
Lambda_250bp_L	5'-CCTCTCAGCCGGGAAAAATGTGCT-3'
Lambda_500bp_L	5'-GGCGCTGTGGCTGATTTTCGATAACC-3'
Lambda_750bp_L	5'-CGGTGAGGAAAAATTTCTGGGT-3'
Lambda_1kb_L	5'-GTTGCCCCGTGAGACAAAGGTACGCC-3'
Lambda_2kb_L	5'-GAACTGACCGCTGAGTCCTATGACG-3'
Lambda_3kb_L	5'-CAGAGGACGCCGGGCCCCGCAGAGC-3'

## ii. DNA Polymerase I sequences

### i. 16S rRNA Sequences:

*Anoxybacillus flavithermus* (T1) JQ267733

*Bacillus stearothermophilus* Y608989

*Carboxydotherrhus ferrireducens* (JW/AS-Y7) U76363

*Escherichia coli* J01859

*Geobacillus kaustophilus* (HTA462) AB002646

*Kosmotoga olearia* NR\_044583

*Marinithermus hydrothermalis* AB079382

*Thermodesulfatator atlanticus* (AT1325) EU435435

*Thermodesulfatator indicus* (CIR29812) AF393376

*Thermodesulfobacterium hydrogeniphilum* AF332514

*Thermosediminibacter oceani* AY703478

*Thermotoga maritima* M21774

*Thermus aquaticus* (YT-1) L09663

### *Geobacillus* sp.M 16S ribosomal RNA gene, partial sequence:

```
1  tcttggtcga ggacgaacgc tggcggcgtg cctaatacat gcaagtcgag
51  cggaccgaac gagagcttgc tcttggtcgg tcagcggcgg acgggtgagt
101 aacacgtggg caacctgccc gcaagaccgg gataactccg ggaaaccgga
151 gctaataccg gataaacacca aagaccgcat ggtctttggt tgaaaggcgg
201 cttcggtcgt cacttgccga tgggcccggc gcgcattagc tagttggtga
251 ggtaacggct caccaaggcg acgatgcgta gccggcctga gagggtgacc
301 ggccacactg ggactgagac acggcccaga ctcctacggg aggcagcagt
351 aggggaatctt ccgcaatgga cgaagtctg acggagcgac gccgcgtgag
401 cgaagaaggc cttcggtcgt taaagctctg ttgtgaggga cgaaggagcg
451 ccgtttgaat aaggcggcgc ggtgacggtg cctcacgaga aagccccggc
501 taactacgtg ccagcagccg cggtaatacg tagggggcga gcgttgtccg
551 gaattattgg gcgtaaagcg cgcgcaggcg gtcccttaag tctgatgtga
601 aagccacagg ctcaaccgtg gagggtcatt ggaaactggg ggacttgagt
651 gcaggagagg agagcggaat tccacgtgta gcggtgaaat gcgtagagat
701 gtggaggaac accagtggcg aaggcggctc tctggcctgt aactgacgct
751 gaggcgcgaa agcgtgggga gcaaacagga ttagataccc tggtagtcca
801 cgcgtaaac gatgagtgtt aagtgttaga ggggtcacac cctttagtgc
851 tgtagctaac gcgataagca ctccgcctgg ggagtacggc cgcaaggctg
901 aaactcaaag gaattgacgg gggcccgcac aagcgggtga gcatgtggtt
951 taattcgaag caacgcgaag aaccttacca ggtcttgaca tcccctgaca
1001 acccaagaga ttgggcgttc ccccttcggg gggacagggg gacaggtggt
1051 gcatggttgt cgtcagctcg tgcgtgaga tgttggttta agtcccgcaa
1101 cgagcgcaac ccttgccctc agttgccagc attcagttgg gcaactctaga
1151 gggactgccg gctaaaagtc ggaggaaggt ggggatgacg tcaaatacatc
1201 atgcccctta tgacctgggc tacacacgtg ctacaatggg cggtagaaaag
1251 ggctgcgaac ccgcgagggg gagcgaatcc caaaaagccg ctctcagttc
```

1301 ggattgcagg ctgcaactcg cctgcatgaa gccggaatcg ctagtaatcg  
 1351 cggatcagca tgccgcggtg aatacgttcc cgggccttgt acacaccgcc  
 1401 cgtcacacca cgagagcttg caacaccgga agtcggtgag gtaaccctta  
 1451 cgggagccag ccgccgaagg tggggcaagt gattggggtg aagtcgtaac  
 1501 aaggtaac

***Geobacillus caldovelox* 16S ribosomal RNA gene, partial sequence:**

1 aagtttgatc ctggctcagg acgaacgctg gcggcgtgcc taatacatgc  
 51 aagtcgagcg ggccaaatcg gagcttgctc tggtttggtc agcggcggac  
 101 gggtgagtaa cacgtgggca acctgcccgc aagaccggga taactccggg  
 151 aaaccggagc taataccgga taacaccgaa gaccgcatgg tctttggttg  
 201 aaaggcggcc tttggctgtc acttgcggtg gggcccgcgg cgcattagct  
 251 agttgggtgag gtaacggctc accaaggcga cnatgcgtag ccggcctgag  
 301 agggtgaccg gccacactgg gactgagaca cggcccacac tcctacggga  
 351 ggcagcagta gggaatcttc cgcaatgggc gaaagcctga cggagcgcag  
 401 ccgcgtgagc gaagaaggcc ttcgggtcgt aaagctctgt tgtgagggac  
 451 gaaggancgc cgttcgaaga gggcggcgcg gtgacggtag ctcacgagga  
 501 agccccggct aactacgtgc cagcagccgc ggtaatacgt anggggcgag  
 551 cgttgtccgg aattattggg cgtaaagcgc gcgcaggcgg tcccttaagt  
 601 ctgatgtgaa agcccacggc tcaaccgtgg agggtcattg gaaactgggg  
 651 gacttgagtg caggagagga gagcgggaatt ccacgtgtag cggtgaaatg  
 701 cgtagagatg tggaggaaca ccagtggcga aggcggctct ctggcctgca  
 751 actgacgtcg aggcgcgaaa gcgtggggag caaacaggat tagataccct  
 801 ggtagtccac gccgtaaacg atgagtgcct agtgttagag gggtcacacc  
 851 ctttagtgct gcagctaacg cgataagcac tccgcctggg gagtacggcc  
 901 gcaaggctga aactcaaagg aattgacggg ggccgcaca agcggtgag  
 951 catgtggttt aattcgaagc aacgcgaaga acctaccag gtcttgacat  
 1001 cccctgacaa cccaagagat tgggcgttcc cccttcgggg ggacaggggtg  
 1051 acaggtggtg catggttgct gtcagctcgt gtcgtgagat gttgggttaa  
 1101 gtcccgcgaac gagcgcaacc ctcgcctcta gttgccagca cgaagtgagg  
 1151 cactctagag ggactgccgg cgacaagtcg gaggaagggtg gggatgacgt  
 1201 caaatcatca tgccccttat gacctgggct acacacgtgc tacaatgggc  
 1251 ggtacaaagg gctgcgaacc cgcgaggggg agcgaatccc aaaaagccgc  
 1301 tctcagttcg gattgcaggc tgcaactcgc ctgcatgaag ccggaatcgc  
 1351 tagtaatcgc ggatcagcat gccgcggtga atacgttccc gggccttgta  
 1401 cacaccgccc gtcacaccac gagagcttgc aacaccgaa gtcggtgcgg  
 1451 taacccttac gggagccagc cg

**Azores\_Furnas isolate 16S ribosomal RNA gene, partial sequence**

1 agcggcggac gggtagtaaa cacgtgggta acctgcctgt aagactggga  
 51 taactccggg aaaccggggc taataccgga tgcttgattg aaccgcatgg  
 101 ttcaatcata aaagggtggt ttgagctacc acttacagat ggacccgcgg  
 151 cgcattagct agttggtgag gtaacggctc accaaggcga cgatgcgtag  
 201 ccgacctgag agggtgatcg gccacactgg gactgagaca cggcccagac  
 251 tcctacggga ggcagcagta gggaatcttc cgcaatggac gaaagtctga  
 301 cggagcaacg ccgcgtgagt gatgaagggt ttcgatcgtg aaaactctgt  
 351 tgtagaggaa gaacaagtac cgttcgaata gggcggtagc ttgacggtag  
 401 ctaaccagaa agccacggct aactacgtgc cagcagccgc ggtaatacgt  
 451 aggtggcaag cgttgtccgg aattattggg cgtaaagcgc gcgcaggcgg  
 501 tttcttaagt ctgatgtgaa agccccggc tcaaccgggg agggtcattg

551 gaaactgggg aacttgagtg cagaagagga gagtggaatt ccacgtgtag  
 601 cggtgaaatg cgtagagatg tggaggaaca ccagtggcga aggcgactct  
 651 ctggtctgta actgacgctg aggcgcgaaa gcgtggggag cgaacaggat  
 701 tagataccct ggtagtccac gccgtaaacg atgagtgcta agtggttagag  
 751 ggtttccgcc ctttagtgct gcagcaaacg cattaagcac tccgcctggg  
 801 gagtacggtc gcaagactga aactcaaagg aattgacggg ggcccgcaca  
 851 agcggtgag catgtggttt aattcgaagc aacgcgaaga accttaccag  
 901 gtcttgacat cctctgacaa ccctagagat agggcttccc cttcgggggc  
 951 agagtgcag gtggtgcatg gttgtcgtca gctcgtgtcg tgagatgttg  
 1001 ggttaagtcc cgcaacgagc gcaacccttg atcttagttg ccagcattca  
 1051 gttgggcact ctaagggtgac tgccggtgac aaaccggagg aaggtgggga  
 1101 tgacgtcaaa tcatcatgcc cttatgacc tgggctacac acgtgctaca  
 1151 atgggcagaa caaagggcag cgaagccgcg aggctaagcc aatcccacaa  
 1201 atctgttctc agttcggatc gcagtctgca actcgactgc gtgaagctgg  
 1251 aatcgctagt aatcgcggat cagcatgccg cgggtgaatac attcccgggc  
 1301 cttgtacaca ccgcccgtca caccacgaga gtttgtaaca cccgaagtcg  
 1351 gtgaggtaac cttttggagc cagccgccga aggtgggaca gatgattggg  
 1401 gtgaagtcgt aacaaggtaa cc

## ii. DNA Polymerase sequences

***Escherichia coli*** DNA Polymerase I (*E.co* Pol I)

GenBank: CP001637

***Thermus aquaticus*** DNA Polymerase I (*T.aq* Pol I)

GenBank: J04639

***Geobacillus stearothermophilus*** DNA Polymerase I (*B.st* Pol I)

GenBank: U33536 (New England Biolabs Ltd.)

***Thermotoga maritima*** DNA Polymerase I (*T.ma* Pol I)

GenBank: AAD36686.1

***Geobacillus kaustophilus*** DNA Polymerase I (*G.ka* Pol I)

GenBank: 06510

***Anoxybacillus flavithermus*** DNA Polymerase I (*A.fl* Pol I)

GenBank:11567

***Kosmotoga olearia*** DNA Polymerase I (*K.ol* Pol I)

GenBank: P001634

***Marinithermus hydrothermalis*** DNA Polymerase I (*M.th* Pol I)

GenBank: CP002630

***Thermosediminibacter oceani*** DNA Polymerase I (*T.oc* Pol I)

Genbank: 9511312

***Bacillus licheniformis*** DNA Polymerase I, ATCC 14580 (*B.li* Pol I)

Genbank CP000002

***Geobacillus sp.M*** LF DNA Polymerase I (*G.me* DNA Pol I),

GeneSys Biotech Ltd. isolate

atc aagaaaaaactcgtcttaattgatggcaacagcgtggcataccgcgcctttttgaccttgccacttttgcataacgacaaaagcattcata  
cgaatgcggttttacgggtttacgatgatgtgaacaaaattttggcggaagaacaaccgaccatttactttagcgtttgacgcgcgaaaaac  
gacgttccggcatgaaacgtttcaagagtataaaaggcgacggcaacaaactccccggaactgtccgagcagtttccgctgttgccgcgagcta  
ttaaagcgtaccgcattcccgcttatgaacttgatcattacgaagcggacgatattatcgggacgctcgcctgcccgcgctgagcaagaagggt  
ttgaagtgaaaatcatttccggcgaccgcgatttaaccagctcgcctcccgtcatgtgacggtcgatattacgaaaaaagggtattaccgacat  
tgagccgtatacaccagagaccgttcacgaaaaatacggcctgactccggagcaaatagtggatttaaaaggattgatggcgataaatccgac  
aacatccccggcgtgcccgcgcatcggggaaaaacggcggtccagctgctgaagcaatttggtagcgtggaaaatgtgctcgcacatcgatcgatg  
aggtgaaaggggaaaaactgaaagaaaacttgcccaacaccgggatttagctctcttgagcaaacagctggcgctccatttgcgcgacgcccc  
ggttgagctgtcgttagatgacatcgtctacgaaggacaagaccgcgaaaaagtcacgcgttatttaagaactcgggtttcagtcgtttcttg  
gaaaaaatggcgcgcgcgga gca gaaggggagaaaccgttgaggagatggagtttgccatcgttgacgtcattaccgaagagatgcttgccg  
acaaggcagcgcgttgctgttgaggtgatggaagaaaactaccacgatccccgattgtcgaatcgactagtgaacgagcatggcgattttt  
tatgccccggagaccgcgctggctgattcgcaatttttagcatggcttgccgatgaaacgaagaaaaaagcatgtttgacgccaaagcgggca  
gtcgttgccctaaagtggaaaggaattgagcttcgcggcgctgcctttgatttattgctcgcctgctcctatttgcctcaatccggctcaagatgccg  
gcgatatcgcgtgcggtggcgaaaaatgaaacaatatgaagcgggtgcggtcggatgaagcggctcatggcaaggt gcc aagcggctcgttccgga  
cgagcagacgcttgctgagcatctcgttcgcaaagcggcagccatttggcgcttgagcagccgtttatggacgatttgcggaataacgaacaa  
gatcaattattaacgaagcttgagcagccgctggcagcgattttggctgaaatggaattcactggggtgaacgtggatacaaacggcgttgaaac  
agatgggttcggagctcgcgaacaactcgtgccatcgagcagcgcatttacgagctagccggccaagagttcaacattaactcgccaaaaaca  
gctcggagtcattttatttgaagctgcagctaccggtgctgaagaagacgaaaacaggctattcgacttcggctgatgtgcttgagaagctt  
gcgcgcgcatcatgaaatcgtcgaaaacattttgcattaccgccagcttggcaaaactgcaatcaacgtatattgaaggattgttgaagttgtgc  
gccctgataccggcgaagtgcatacgaatgttcaaccaagcgtgacgcaaaactggcggtcagctcggccgagccgaacttgcaaaacattcc  
gattcggctcgaagagggggcgaaaaatccgccaaagcgttcgtcccgtcagagccggactggctcattttcgccgcgagattactcacaattgaa  
ttgcgcgtcctcgcccatatcgccgatgacgacaatctaattgaagcgttccaacgcgatttggatattcacacaaaaacggcgatggacattt  
tccatgtgagcgaagagggaagtcacggccaacatgcgcgcgagcgaagggcgttaacttcggtatcgtttacggaattagcgattacggatt  
ggcgcaaaacttgaacattacgcgcaaggaagctgccgaatttatcgaaacttacttcgccagctttccgggcgtaaagcagtatatggaagaaac  
attgtgcaagaagcgaacagaaaggatatgtgacaacgctgttgcatcggcgccgctatttgctgatattacaagccgcaatttcaacgtcc  
gcagttttgcagagcggacggccatgaacacgcccattcaaggaagcgcgctgacattatataaaagcgatgattgatttagcggcacggct  
gaaagaagagcagcttcaggctcgtcttttgcgtcaagtgcagcagctcattttggaagcgccaaaagaggaaattgagcgattatgtgag  
cttgctccggaagtgatggagcagggcgttacgctccgcgtgccgctgaaagtcgactaccattacggcccaacatggatgatgccaac cga

Translation of DNA sequence:

██████klvlidgnsavayraffalpllhndkgihtnavygftmmlnkilaeeqpthllvafdagkttfrhetfqeykggrqqtppelseqfpllrel  
lkayripayeldhyeaddiigtlaaraeegfevkiisgdrdlqtqlasrhvtditkkgitdiepytpetvhekygltpedivdlkglmgdksd  
npgvpgigektavqlkqfgtvenvlasidevkgeklkenlrqhrdlallskqlasicrdapvelslddivyegqdrekvialfkelgfsfl  
ekmaapa██████egekpleemefaivdviteemladkaalvvevmeenyhdapivgialvnehgrffmrpetaladsqflawladetkkksmfakra  
vvalkwkgielrgvafdlillaayllnpaqdagdiaavakmkqyeavrseavygkg██████krsldpdeqtaehlvrkaaiwaleqpfmddlrneq  
dqltkleqplaaailaemefgtgvnvdtkrleqmgseqlaeqlraieqriyelagqefninspkqlgvilfeklqlpvlkktktgystsadvekl  
aphheivenilhyrqlgklqstyieglkvvprdtgkvhtmfngaltqtgrlssaepnlqnipirleegrkirqafvpsedwlfifaadysqiel  
lrvlahiaddnlieafqrdldihtktamdifhvseevtanmrrqakavnfgivygisdyglaqlnitrkeaaefieryfasfpgvkqymen  
ivgeakqkgvyttllhrrrylpditsrnfvrfsaertamntpiqgsaadiikkamidlaarlkeeqlqarlllqvhdelileapkeeerlce  
lvpevmeqavtlrvplkvdyhygptwydak

***Geobacillus caldovelox* LF DNA Polymerase I (G.ca Pol I),**

GeneSys Biotech Ltd. Isolate

atggaagaggaaaaaccgctggccaagatggcatttacgcttgctgaccgcgtgacggagagatgcttgcgcacaaggcggcgcttgctgcttg  
aagtggctcgaggaaaattatcatgatgcgccgatcgctcgccatcgctgtggtcaacgaacatggacgggttttctctgcgccggagacggcgct  
tgccgatccgcagtttgctgcctggcttggtgatgaaacgaagaaaaaagcatgtttgactcaaagcgcggcgagtcgccttgaaatggaaa  
ggaattgagctatcgggcgcttctcttgatttattgctggccgctatttgcttgatccggcgcaaggtgttgatgatgtggctgccgcagcaa  
aaatgaagcaatacgaagcggcgccgcggcgatgaagcgggtgatggcaaaagggcggaagcgggctgcccgatgagccagtgctcgccgagca  
tttggtccgcaaggcggcgcgatttggcgctcgaacgtccgtttttggatgagctgcgccgcaacgaacaagatcggttgctcgctcgagctc  
gagcagccgttgctctcgattttggcggaatggaatttgccggagtgaaagtgatacgaagcggctcgaacagatggcggaagagctcgccg  
agcagctgcccacggtcgagcagcgcatattatgagctgcgccgccaagaattcaacatcaattcatcgaaaacggctcgcgctcattttattga  
aaaactgcggctgcccgtcttgaaaaaaacgaaaaccggtactccacttcggcggtatgtgcttgaaaaacttgccgcttatcacgagatcggtg  
gaaaacattttgcatttccgcagcttgccaagtgcagtcgacgtatattgaaggattgctgaaagtcgtgacccgatacaaagaaggtgc  
atacgaatttcaatcaggcggtgacgcaaacggagcggctcagctcgacggagccgaacttgcaaacattccgattcggttgaggaaggacg  
gaaaatccgccaagcggttcgtgccgtcgaggtctgattggctcattttcgctgccgactactcgcaaatgagttgcgctcctcgcccatatt  
gcggaagatgacaatttaagtgaagcggtccgcccgcgatttggatatccatacgaaaaacagcgatggacattttccaagtgcgagcaggacgaag  
tgacgcccacatgcgccgtcaggcgaaggcggtcaactttgggatcggttacgggatcagtgattacggcttgccgcaaaaacttaaatatttc  
acgcaaaaggccgctgaattcatcgagcgctacttcgaaagcttccctggcgtaagcgggtatatggaaaaacttgtaagaagcaaaaacag  
aaagggtatgtgacgacgtgctgcacggcgccgctatttgccggatatcacgagccgcaacttcaacgtccgcagctttgctgaacggatgg  
cgatgaacacgcccattcaaggagcgccgctgacattattaaaaaggcgatgatcgatctgaacgccagactgaaggaagagcggctgcaagc  
gcgccttttgctgcaggtgatgacgagctcattttggaggcgccgaaagaagagatggagcggctgtgccggctcggtccggaagtgatggag  
caagcgggtcacacttcgctgcccgtcaaagtcgattaccattacggctcgacatggtatgacgcgaaa██████aaa

Translation of DNA sequence:

██████eeekplakmaftladrvteemladkaalvvevveenyhdapivgiavvnehgrfflrpetaladpqfvawlgdetkkksmfdsakraavalkwk  
gielcgvsfdlllaaylldpaqgvddvaakmkqyeaarpeavygkgakravdepvlaehlvrkaaiwalerpflldelrrneqdrllvel  
eqplssilaemefagvkvdtkrleqmgelaqlptveqriyelagqefninskrllgvilfeklrlpvlkktktgystsadveklapyheiv  
enilhfrqlgklqstyieglkvvprdtkkvhtifngaltqtgrlssstepnlqnipirleegrkirqafvpsedwlfifaadysqielrvlahi  
aeddnlmeafrrldihtktamdifqvsedevtpnmrrqakavnfgivygisdyglaqlnlsrkeaaefieryfesfpgvkrymenivgeakq  
kgvyttllhrrrylpditsrnfvrfsaermamntpiqgsaadiikkamidlnarlkeerlqarlllqvhdelileapkeemerlclrvpevme  
qavtlrvplkvdyhygstwydak

***Carboxydothemus ferrireducens*** LF DNA Polymerase I (C.fe Pol I),  
DSMZ 11255

atgaagaattgaacccaaaaataaagaagaataaccaggaaggtaaagatttgggtgcaatttgaaactgtagaaacggaaggacagattgcag  
tagtttttagtgatggattttacgttgatgacggggaaaaaacaagttttactcgtagaccggtgaatgaaatacaggaaatatttaggaa  
taaaaaattattaccgacgatgccaagaatttatcatgtctgtttagaaaaaggtctgactttcccgaaagtttggttgatgcgcggatt  
gcagcttatgttttaacccggccgacccaaatcccggtcctcaaggggtttatctaaagtatgacttaccgggtgatgaagatgtatctttaa  
acattagagggttgttttatttaaaaaaagaatgatgaaaaaatctttgagcaggagcaggaaaggttattttatgaaatagaacttccttt  
aactccagttccttctcaaatggagcataccggcattcaggttgaccgggaagctttaaagagatgtcgttgagctgggagagcaaattgaa  
gagttaatccgggaattttatgcgtggcggggaagagttaacttaaatcgcaccaagcagctgggagttattcttttgaaaaacttgggc  
tgccggtaattaaaaagaccaagacgggtactctaccgatgcggaggttttgaagagctcttgccctaccacgaaattatcggcaaaatatt  
gaattaccggcagcttatgaagttaaaatccacttatactgacggcttaatgccattaataaatgagcgtaccggttaaacttcacactactttt  
aaccagaccgggtactttaaccggacgcctggcgtcttcggagcccaatctccaaaatattcccatccgggtggaactcggtcggaattacgca  
agatgtttatacccttcaccggggatgattatattgtttcggcggattattcccagattgaattaaggcttcttgccatttttccgaagagcc  
caagcttattgaagcttaccaaaaaggggaggatattcaccggaaaacagcctccgaggtgttcgggtgatcttttgaagaagttactcccgag  
atgcgcgctcatgccaaagtcggtgaacttcggcattgtttatggcattagtattttgttttaggcagagacttaagattccccgggaggttg  
ccggttaagtacattaaaaattattttgccaactatcccaaagtgcgggagtatctcgacgaacttgccgtacggcaagagaaaagggttatgt  
gaccactttatttggcgaaagacgctatattcctgagctatcttcaaaaaaccgcacggttcaggggttggcgaaaggacggcgatgaatact  
ccccttcagggtcgggtgccgatattattaagcttgcaatgattaatgtagaaaaagaacttaagcccgtaagcttaagtcccggtccttc  
tttcgggtgcacgatgagtttagttttagaagtgccggcggaagagctggaagaggttaaagcgctggttaaaggggttatggagtcggtggttga  
actgaaagtgcctttaatcgctgaagttggtgcaggcaaaaactggtatgaagcgaagtaa

Translation of DNA sequence:

kieiepkikkeyqegkdlvqfetvetegqiavvdsdgyvddgektkfysldrlneiqrnkkiitddakgiyhvclekgltfpevcfdari  
aayvlnpadqnpnglkglylkydlpvyedvslnirglfylkkemmkkifeqeqlfyielpltpvlaqmehtgigvdrealkemslelgeqie  
elireiyalageefnlnspkqlgvilfeklglpvikktktgystdaevleellpyheiiqkilnyrqlmklkstytdgmlplinertgklhttf  
nqtgtltgrlassepnlqnipirlelgrklrkmfipspgydyivsadyseielrlahfseepk lieayqkgedihrtasevfsvsleevtpe  
mrahaksvnfgivygisdflgrdlkiprevagkyiknyfanypkvreyldelvrtarekgyvttlfgrrryipelssknrtvqgfgertamnt  
plqgsaadiiklaminvekelkarklksrlllsvhdelvlevpaeleevkalvkgvmesvvelkvpliaevgagknwyek

***Thermodesulfobacterium hydrogeniphilum*** DNA Polymerase I (T.hy Pol I),  
DSMZ 14290

atgatatattcttattgatggatcttcttttgtttatagagcctattttgctattcctggatatttagctactactaagggtatttctactaaag  
ctattttttgggtgttactcaaatgatcctaaaaattctaaaaagaatggggtccagaatatattatctggtttatggatgaaaaagaacctacttt  
tagacatatagcctatgaaaattataaagctaccggtccaggaatgccagaggatctgaaaatccagattccttatataaaaaaattatacca  
gccctcggaataacctgttatttctgccccgggatagaggcagatgatcttattgcaacttttgtccataaaaataataaaaaattagataaat  
cggctattattgtggcaggagataaagatctttattctctgttttcagaaaaggtttgcatttatgatcctgttagagagaaacttttagataa  
agaagcttttttaaaaaataggttttgcctactatttttcagaaattccgtgctttgaccggagataaaagtataatattcctggtgtc  
ccgggagtgaggagaaaaaactgctaaaaatcttattggaaaatttaagaatctgaataatctttatgctaataattcaaaactaaccctttaa  
aacttagagaaaaacttttaaaatacaaaagaacaggttttaataatgtagctctccttactcttaatttttaatgctccctcccttcagaaaag  
tcttgaatactacgaaagaaaagatccggattacaattttctaaaaaactttttaaagaattggaatttagaataatttaacagaattaaaa  
attccttctgcgaacttaagcaagattagttgaagaaaaatagatcttgaaaagcttaaaaaagcagaaattttaagtattaaattcagg  
gaaacctcttttcttaaccaatcctgaagaattaacagttgcctctgatgaaaagagggttataaaactttctgtcgaagtttttaaaatct  
aaagaaaaatgctaagtctgaatttattcttcatgatcataaaagttttctaaaaaattacgaaatttttttagattttctttttgataccaaa  
ttagcagcttatcttctaatccatctcttaagaattatgatttagattttcttttgcaggagcatttagatattaatttaaaactctttaaaag  
cctctaaaaatgaactgcaggctataaaaacatcggggtttatttggtaggaaggatttaagaaaaaagtggaagaagagaatttaaccac



ttggccttaaaaaagtagagattccttttgagtgaagtcctttttgagatggaaaaaaggggtttaaaattgacatagaatatgtaagagaactt  
aaccagagatatcagaaaaagactaaaagaatagaagaaaagcctttttgaaatagcaggttttagatttaattcctcgttccagtagagaggtag  
gagttattctttttgaaaaacttaaacctccctatgataaaaaaacacccaaaagttctttaccctctacagatgcagaagttttggaagaact  
tgctcctcttcatccctttgtaaagctccttctcaatataggactttatataaaaaataaaagcacttatttagaagccttttttaaaatagt  
tcttcttctgattggcggtttacatactgaattcaatcaaaactggaacagccacagggaggtctttgttctcaaaatcctaatttacaaaatattc  
ctataaaaaggtgatgaaggattagctataagaagggtttttatctcagaagaagggtatttactgtgtagtttggtatttcacaaatcgaatt  
aagaattcttgctcatttctcagaagatgagaatcttatttagagcttttgaaagaggggaggatattcatacttttactgctgtgaggttttt  
ggtgtgcccccttgaaaaggttactccagaaatgagaagaatgagtaaggctattaattttggcatagcctatggaatgagtgcttatggtttta  
gtaagaattaagaataagtgtaaaagaagcagaggtctataattaacagatatttttagcagatatcctaaaattaagagtatatagagaaaac  
tatagaatttgccagagaaaatggttatgtgaaaactctggcaggaagaaagcgttatattcctgaaatatttagtccataaaaactgtaaag  
gagctggggcagagaatagcagtaaatcacctattcagggtcagcctctgatcttataaagtgtgctatggtggtcctttataaagaactta  
aaaaacaaaatctaaaaactctattatattacaaattcatgatgagttgatcttcgaggttcctaaggaggaagtcgaaattgtaaaagaagt  
tgcacctaaaataatggaaagtcctataaaactcaaagttcctattaaagttaatatgaccatttggtaaaaactgggcagaa **taa**

Translation of DNA sequence:

**M**iflidgssfvryrayfaipgylattkgtfptkaifgvtqmilkilkewgpeyilwfmdekeptfrhiayenykatrpgmpedlkiqipyikkiip  
alqipviscpgyeaddliatfvhkiikkldksaiivagdkdlyslvsekvciydpvreklldeaffkkygfdptifefraltgdksdnpgv  
pgvgektaknligkfklnlnnlyaniskltplklrenllkykeqvlennalltlnfnaplpsesleyyerkdpydnflklfkelefr **K**ftelk  
ipsaelkarlveekidleklkkaeilsikiqgnlfsltneeltvasdekeayklsvevfknlkknaksefilhdhksflknyefldfldtk  
laayllnpslknydlldflqehldinlnslkasknelqaiktclgyllylgdklkkveeenlwtlkkveiplsevlfemekkgfkidieyvrel  
nqryqkrlkeieeklfeiaqfrfnprssrevgvilfeklklpmikktpksslpstdaevleelaplhpfvkllyqyrtlykikstyleaflkyv  
sssdwrlhtefnqtgtatgrlcsqnpnlqipkgdeglairrvfiseegyllcsl **dysqiel**rilahfsedenlirafergedihtftacevf  
gvplekvtpemrrmskainfgiaygmsayglskelrisvkeaeaiinryfsrypkikeyiektiefarengyvktlagrkryipeifspnktvk  
elgqriavntpiqgsasdlkamvalykelkkqnktsiil **qihdel**ifevpkeeveivkevapkimespyklvpikvnmmtigknwae

***Thermodesulfatator indicus*** DNA Polymerase I (*T.in* Pol I),

DSMZ 15286

gcatacnagtttatccgcatcaaaaangactttggcttcaagggttttagccgcactttgntccgaaaacgatgggtggctatggccttcagct  
acctttttcaatgaaatcagccggcaatttgtaanccggcaagtattttcntttacctgttcagccccgtaagcggcggtggcacacttggcct  
ttggtgcgctttttcagccggcctatatcgtgaaaaagcgcggctactttaaccacgtcaaggtctgctcctctgccttggccaggtgtt  
cagccaggcgtaaaacgcgcattacgtgtgtaaacgatggcttccgcgctttgcctcgtaaaaggcgttcagcttcccttagcaaatTTTTCCA  
gagtttttgcgtttacgattcccactaaaaagattagtaacaagagagaaacaaatttgcaaatgggttaggaagttt **at**ggctcaaaaagtttgt  
ttcctaaaaaattaccatttaaagatgataaagaccccatcttcgttattgacgggagttcttttgtttaccgggcttactatgccataagagg  
gcatactatcaaacgcgaagggtcccaaccaaggcgtctttgggtttaccagatgcttttaagccttttgcgtgagatgaacctgagtat  
gtggtggtgtgctttgacgcaaaaggcctacttttcgccacgagatgtacaagaatacaaaagccaaccgccaccatgccagatgatcttt  
ccgtccagattccctatatcaaaagagtaaccagggcctttggagtcctattcttgaaatagaaggctttgaagctgacgatctcatcgccgc  
tattgccactcgtatggaagaccaattgtcatcgttggtggagataaaagatttgttcccccttatttcagagaaagttgtcatgtgggacccc  
atgaaagacgaactgattgacgaaagctggataaagaaacgttttggcattgaacctaaaaagctccttgatgtaagggcccttgccggcgata  
gcattgataacgtgccaggggttccgggtattggtgaaaaaacggccctaaggctcataaaagaatacggttcccttgaaagaagtccttaacca  
tgccgaagaataaaaacaaaagcgttgctgtgaaaacctcatcaaacacgcggagacgccttattttcaaaaaactggttgagcttgaggcc  
aaagcccaatcccccttgagcctgatttttaccgcaaacggccattaaatgccctaaaactaagggaactcttccttgagcttgaatt **aaq**a  
agctcttaaaaggaacttccagctactaaaaccctttcgtatgaccagtagcagctggttcttgacccggataaagtaaaagaaattgtagaaaa  
ggccaaaggggcccgaagtgggtggtatt **gac**tt **gaa**agtatacgaagaccccatgcgtgggaaaatagtaggggtctcgctttgttttaac  
ccgccccaaagcctattatttcccttttagacatgaaggccttgaggcccaaaagcagcttccctgggaggccttactcatctggccagcctca  
ttgaagacccctcagttaaaaagataggccacaatatcaagtatgacttgattattcttgcctgcctacggcgtaactttaaggggccttgagg  
ggataccatgctggtctcgtatctccttgatccaacacgtcgtacccacggccttgatgagctggccgaagaggtcctggggcataccatgatt  
tcttacaaggaagtgactaaagaactggccaaaggagagagctttgccaggggtccctcttgaaaaggcaaaagtttacgcctgtgaagacgccc  
acgttacctatctgctttatcaatatcttggcccaactcaaagaggaaagcctctggaaggtctttacggaaattgatcgacctttaataga

agttttggccacatggaatggtaggtattaagattgacaccgcctatcttagaggactttcgcgagaaatggctgaaaagttaaaggagctt  
gaagaaaaaatttacacctggctggtgaaaaatttaatatcaattccagcaacaactgggccagattttatttgaaaaactaaaactcccta  
cggtaaaaaagacccccaaaaaacggcctattcaacggataacgaagtattagaggaaactttctgcggtccacgaacttccgcgtctgatact  
tgagtatagaactctggctaaactcaaatctacttatgttgatgcctcccgaaagatgggtaaatcctgaaactggctgtcttcatacttccctt  
aaccagacggttacggccactggaagactttcaagcagtgaacctaatcttcaaaatattcctgtgcgtggtgaagaggggcttaagattcgcc  
aggcctttgtgcccggaggagggtttcttttctctcagctgactattctcagattgatctgcgagttttagcccattactcgggagatgaaac  
cttgattaaggccttctggcagggggaagacattcacggcgacggctgcgaaaattttggtatccccgccagaagaagtaactcctgagatg  
cggcgtatggccaagactataaactttggcattgtttacggcatgagtccttacggtctgcgaaagaactcaaaatggccgacctgaggcca  
aggcctttattgagcgctattttgaacgctaccagggtgtgaaacgctatatggaacaaatcgtggctgaagcccagaaaaagggtacgtgga  
gaccttttgcgacgcaaaaggcctcttctgacatcaatagccctaatactgacggcgcgagtttgccgagcgacggctataaacactcct  
attcaggggacagccgctgatattatcaagctcgcatgataaaaattcacgggatttttaagaaaaaggctttgggacaaggatgcttcttc  
aggtgcacgacgaacttcttttgaagtgcctgaaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtggttgaatt  
gaaggttcctctaaaagtaaacctggcaatagggaataatgggcagaggcaaggca**tgga**

Translation of DNA sequence:

**l**aqkslfpkklpfdkddkpifvidgssfvrayyairghlsnrkglpkavfgftqmlklkllremnpeyvvcfdakgptfrhemykeykanrp  
pmpddlsvipyikevtrafgvpileiegfeaddliaaiatrmerpivivgdkdlfplisekvvmwdpmkdelideswikkrfgiepkklldv  
ralagdsidnvpvgvgigektalrlikeygsleevlnhaeekqkrlrenlikhagdalisksklveleakapiplepdyfrkrplnalkrelf  
lelefk**l**llkelpatktlsydqyelvldpdkvkeivekakgaevaid**l**esdtkdpmrgkivgvsfcfnppkayyfpfrhegleaqqlpweaf  
thlasliedpsvkkighnikydliilarygvtlkglegdtmlasylldprrthgldeaevlghtmisyevtkelakgesfarvplekakv  
yacedahvtyllyqyfwpkleeslwkvfteidrplievlahmemvgikidtaylrglsremaeklleekiyltagefkninsskqlgqilf  
eklklptvkktpkktaystdnevleelsavhelprlileyrtlaklkstyvdalpkmvnpetgrlhtsfngtvtatgrlsssdplnqnipvrge  
eglkirqafvpeegflflsa**dysqidl**rvlahysgdetlikafwggedihrrtaaeifgippeevtpemrrmaktinfgivymgspyglakelk  
igrreakafieryferypgvkrymeqivaarekyvetlfgrkrplpdinspnrtarefaertaintpiqgtaadiiklamikihrifkekgf  
gtrmll**qvhdell**fevpekeieeiqpivrqimegvvelkvplkvnlaigknwaeaka

***Thermodesulfatator atlanticus*** DNA Polymerase I (*T.at* Pol I),

DSMZ 21156

**atc**gcacaaaaaatctatttccccaaaaattaccctttgacggcaaaaaagaccctatttttgtgattgacggaagctcttttatttatcgcg  
cctattttgctattaaaggccacttatcaaacgtaaaagggtcccgaccaaggctatttttggttttaccagatgcttctcaaactcctgaa  
agaaatggacccaaaatacgtggtggtctgttttgatgccaaaggccctacttttcgacacaaagtttacgaggaatacaaaagcaaacgctccg  
gccatgcctgatgatttggccgtgcaaatcccttatatccggaagtccaccagggtccttggcgctacctatccttgaaatagaagccttgaag  
cggacgacttaattgctgccattgccaccgcattgaccatcccattgtcatcgttggaggggacaaagatctatttcccttatctctgaaaa  
agtcgtcatgtgggaccgatgaaggacctattcattgaccaaacctggattaaagaacgttttggggttgagccgaaaaactcttgatgta  
agggcccttgccggagacagcatcgataacatccccgggttccaggatttggcgaaaaaactgccttaaaactcatcaaagagtacggctctc  
ttgaagaagttttaagcacgctcacgagatcaacaaaaacgcttgcgcgaaaaacttataaagtatgccgaacaagccgctctctcaaaaa  
attagtacagcttgccaaagaggccctattcccttgagccagattttttcaggcgccgcgagccaaactcctcaaattacgagagctattc  
cttgagcttgagttcaaa**aaa**cttctcaaaagaactccccgccacaaaaccatttcttatgacgaatacgaactcgtgacagattcttcccgaa  
tgcacgaaattttagccaaagctcgtgaaaaggcctggtggttatt**gac**ctt**gaa**agtaatcacatagacccccatgcgcgggaaaaatagtagg  
agtggccctttgttttgaaaccctaaagcttattattttcccttgcgcacgaaggcttagaggccaggaagcaacttccctgggaagccttt  
ggcgacatagcagccctcatcgaggaccaacaagtcaaaaaatcgccacacataaaatagcactgattcttctggcagcgtatggagtag  
ccttgaaagggttgaaggtgataccttgcctgcgtctatcttttaaacctaccagacgcactcatggccttgatgaattagcagaagaat  
cctcggctcactgcatgatctctacaaagaagtcacaaaagaacttgccaaaggagaaagtttcgcaagagtccacttgaaaaagccgaagac  
tatgctgtgaagacgcccacgtaacctatctctttatcaatacttctggcccaattaaaagaagaaagtctctggcgagtttttgaaagaga  
tagaacgtcctttgatcaagtgctggctcgcatggaaatggcaggcatcaaaatagacgtaccttatctgcgagcgctttcccaagagcttgc  
ccaaaaactaaaagaacttgaacaaaaaatctacgaaatagcccaggaacaattcaacataaaactcaagtaggcagcttgcctatcctttt  
gaaaaacttaaaactccccaaagttaaaaaaacccctaaaaagacggccttttccacagacaacgaggttttagaggagctttccactcttcatg  
aacttccgcgcctggttcttgaaatcgcacgttggccaagctcaaatcaacttacgtagatgccctcccaagatggctaatacctgagaccgg

ccgcattcatacttcctttaaccaaaccatcactgccacgggaaggctctcaagcagtgatcccaatctccagaatattcctgtacgcgcgaa  
gaaggcactaaaattcgcaaggcctttgtcccggaaaaagggctctttttctttcagcagattattccagatagacctccgggtacttgccc  
attattcaggcgatgaaacattgatagaagcctttaagcgcggaagatattcaccgcccagcagagatctttggtgtttccctga  
agaagtcacctctgagatcgggcgcatggccaaaaccataaacttcggcattgtttacggcatgagcccttacgggctagccaaagaactaaa  
attggccgtcgcgaggccaaagcctttattgagcgttattttgaacgctaccccggtgttaaagcgtacatggaacaaattgtcgctgaagcca  
gagaaaaaggctatgtggaacacttttcggacgtaaaaggcctttgcccgatattaacagcccaatcgtagcgacgtgagtttgccgaacg  
cactgccatcaataccccattcaaggaaccgcgctgatcatcaaaacttgccatgatcaagctagacaccactattgaagaaaaaggcttt  
gagaccaagatgcttcttcaagtcattgatgaactcctttttgaagttccagaaaaagaagttgaagaaattcaaaaaattgtccgacaaatca  
tgaaggagttgtgactttaaaagtcctttaaagtaaaccttgctttagggaataattgggcagaggcaaggca **taa**

Translation of DNA sequence:

**l**aqknlfpqklpfdgkdpifvidgssfiyrayfaikghlsnrkglptkaifgftqmlklkemdpyvvvcfdakgptfrhkvyeykanrp  
ampddlavqipyirevtrafgvpileiegfeaddliaaiatridhpivivggdkdlfplisekvvmwdpmdlfdqtwikerfgevpeklldv  
ralagdsidnipgvpigiektalklikeygsleevlkhaheikqkrlrenlikyaeqarlskklvqlakeapiplepddfrraanvklrelf  
lelefklkelpatktisydeyelvtdssrmheilakarekglvvidl **l**snhidpmrgkivgvalcfepkayyfpfrheglearkqlpweaf  
gdiaaaliqdqqvkkighnikydillarygvalkglegdtllasyllnptrrthgldeaeilghcmisykevtkelakgesfarvplekaed  
yacedahvtyllyqyfwpkleeslwrvfseeierplikvlarmemagikidvpylralsqelaqklkeleqkiyeiaeqfninssrqlahilf  
eklklpkvkktpkktafstdnevleelstlhelprlvleyrtlaklkstyvdalpkmanpetgrihtsfngtitatgrlsssdplnqnipvrge  
egtirkafvpekgalfls **adysqidl**rvlahysgdetlieafkrgedihrrtaaeifgvspeevtsemrrmaktinfgivygmspyglakelk  
igrreakafieryferypgvkrymeqivaearekgyvetlfgkrplpdinspnrtarefaertaintpiqgtaadiiklamikldttieekgf  
etkml **qvhdell**fevpekeveeiqkivrqimegvvtlkvplkvnalalgnwaeaka

### III. DNA Polymerase I chimera sequences

#### i. *T.in-G.me\** Pol I

**TG\_C1 :**

TGC1_U1	5'-tctcagctgactattctcag <b>attgaattgcgcgtcctcgccca</b> -3'
TGC1_L1	5'-tgggcgaggacgcgcaattcaatctgagaatagtcagctgaga-3'
TGC1_U2	5'- <b>gttaacttcggtatcgtttacgga</b> atgagtccttacggtctggc-3'
TGC1_L2	5'-gccagaccgtaaggactcattccgtaaacgataccgaagttaac-3'

**atg**ggcctcttaaaggaacttcagctactaaaaccctttcgatgaccagatacagagctggttcttgaccggataaagtaaaagaaattgtag  
aaaaggccaaagggccgaagtgggtggtattgaccttgaaagtatacgaagaccccatgctgggaaaaatagtaggggtctcgctttgttt  
taaccgcgccaaagcctattatttcccttttagacatgaaggccttgaggcccaaaagcagcttccctgggaggcctttactcatctggccagc  
ctcattgaagaccctcagttaaaaagataggccacaatatcaagtatgacttgattattctgtcgctacggcgtaactttaaagggccttg  
aaggggataccatgctggcttctgatctccttgatccaacacgctcgtaaccacggccttgatgagctggccgaagaggctctgggcataccat  
gatttcttacaaggaagtactaaagaactggccaaaggagagagctttgccagggtccctcttgaaaaggcaaaagtttacgcctgtgaagac  
gccacgcttacctatctgctttatcaatatcttgcccaaaactcaaagaggaaagcctctggaaggcttttacggaaattgatcgaccttaa  
tagaagttttggccacatggaaatggttaggtattaagattgacaccgcctatcttagaggactttcgcgagaaatggctgaaaagttaaagga  
gcttgaagaaaaatttacaccctggctggtgaaaaatttaatatcaattccagcaaaactgggcagatttttatgtgaaaagctaaaactc  
cctacggttaaaaaagaccccaaaaaaacggcctattcaacggataacgaagtattagaggaaactttctgcggtccacgaacttccgcgtctga  
tacttgatgatagaactctggctaaactcaaacttacttatgttgatgcctcccgaaagatgggtaactctgaaactggctcttcatacttc  
ctttaaccagacggttacggccactggaagactttcaagcagtgaccctaactcttcaaaaatattcctgtgcgtggtgaagaggggcttaagatt

cgccaggcctttgtgccggaggagggtttcttttctctcagctgactattctcagattgaattgcgcgctcctcgcccatatcgccgatgacg  
acaatctaattgaagcgttccaacgcgatttggatattcacacaaaaacggcgatggacattttccatgtgagcgaagagggaagtacgggcaa  
catgcgcccgccaggcaaaaggcgttaacttcggtatcggttacggaatgagtccttacggtctggcgaaagaactcaaaattggccgccgtga  
ggccaaggcctttattgagcgctattttgaacgctaccaggtgtgaaacgctatatggaacaaatcggtgctgaagcccgagaaaagggtac  
gtggagacccttttcggacgcaaaaggcctcttcctgacatcaatagccctaactcgtagcgcgcgagtttgccgagcgacggctataaaca  
ctcctattcaggggacaagcgctgatattatcaagctcgccatgataaaaattcaccggatttttaagaaaaaggctttgggacaaggatgct  
tcttcaggtgcatgacgagcttattttgaagcgccaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtgggtgaa  
ttgaaggttctctctaaaagtaaactggcaatagggaaaaattgggcagaggcaaggcaaaa

Translated protein sequence:

gllkelpatktlsmtryelvldpdkvkeivekakgaevvaidlesdtkdpmrgkivgvsclfnppkayyfpfrhegleaqkqlpweafthlas  
liedpsvkkighnikydliilarygvtlkglegdtmlasyllldptrrthgldeelaevlghtmisyevtkelakgesfarvplekakvyaced  
ahvtyllyqyfwpkleeslwkvfteidrplievlahmemvgikidtaylrslremaekleekiylagekfninsskqlgqilfeklkl  
ptvkktpkktaystdnevleelsavhelprlileyrtlaklkstyvdalpkmvnpetgrlhtsfngtvtatgrlsssdnlqnipvrgeeglki  
rqafvpeegflflsadyseelrvlahiaddnlieafqrdldihtktamdfhvseeevtanmrrqakavnfgivymspyglakelkigre  
akafierierypgvkrymeqivaarekgyvetlfgkrplpdinsprntarefaertaintpiqgtaadiiklamikihrifkekgfgtrml  
lqvhdelifaeapkeieeiqpivrqimegvvelkvplkvnlaigknwaeaka

## TG\_C2:

TGC2\_U1 5'-ctataaactttggcattgtttacggaattagcgattacggattggc-3'  
TGC2\_L1 5'-Gccaatccgtaatcgctaattccgtaaaacaatgccaaagtttatag-3'  
TGC2\_U2 5'-tgaacacgcccgattcaaggaagcggcgctgatattatcaagctcgccat-3'  
TGC2\_L2 5'-atggcgagcttgataaatatcagcggcgcttccttgaaatcggcgtgttca-3'

atgggctcttaaggaacttcagctactaaaaccccttcgatgaccagatacgagctgggttcttgaccgggataaagtaaaagaaattgtag  
aaaaggccaaaggggcggaagtgggtggtattgaccttgaaagtatacgaagaccccatgctgggaaaaatagtaggggtctcgctttgtt  
taaccgcgcaaaagcctattatttcccttttagacatgaaggccttgaggccaaaagcagcttcctgggaggcctttactcatctggccagc  
ctcattgaagaccctcagttaaaaagataggccacaatatcaagtatgacttgattattcttgctcgctacggcgtaactttaagggccttg  
aaggggataccatgctggttcgtatctccttgatccaacacgctcgtaaccacggccttgatgagctggccgaagaggctcctggggcataccat  
gatttcttacaaggaagtgactaaagaactggccaaaggagagagctttgccagggtccctccttgaaaaggcaaaagtttacgcctgtgaagac  
gcccacgttacctatctgctttatcaatatttctggccaaaactcaaagaggaaagcctctggaaggcttttacggaattgatcgaccttaa  
tagaagttttggccacatggaaatggtaggtattaagattgacaccgctatcttagaggactttcgcgagaaatggctgaaaagttaaagga  
gcttgaagaaaaatttacaccctggctgggtgaaaaatttaatatcaattccagcaaaactgggccagattttatgtgaaaagctaaaactc  
cctacgggttaaaaagacccccaaaaaaacggcctattcaacggataacgaagtattagaggaactttctgcggtccacgaacttcgcgctga  
tacttgagtatagaactctggctaaactcaaactctacttatgttgatgcctcccgaaagtggtaaatcctgaaactggctcttcatacttc  
ctttaaccagacggttacggccactggaagactttcaagcagtgaccctaactctcaaaatattcctgtgcgtgggtgaagaggggcttaagatt  
cgccaggcctttgtgccggaggagggtttcttttctctcagctgactattctcagattgatctgcgagttttagcccattactcgggagatg  
aaaccttgattaaggcctcttggcagggggaagacattcaccggcgacggctgcagaaaatttttggtatcccgccagaagaagtaactcctga  
gatggcggtatggccaagactataaactttggcattgtttacgggaattagcgattacggattggcgcaaaacttgaacattacgcgcgaagga  
agctgccgaatttatcgaacgttacttcgccagctttccgggctgaaagcagtatatggaacacattgtgcaagaagcgaacagaaaggatat  
gtgacaacgctgttgcatcggcgcgctatttgctgatattacaagcgcaatttcaacgtccgcagttttgcagagcggaacggccaagaac  
acgcccattcaaggaagcggcgctgatattatcaagctcgccatgataaaaaattcaccggatttttaagaaaaaggctttgggacaaggatg  
cttcttcaggtgcatgacgagcttattttgaagcgccaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtgggtg  
aattgaaggttctctctaaaagtaaactggcaatagggaaaaattgggcagaggcaaggcaaaa

Translated protein sequence:

mgllkelpatktlsmtryelvldpdkvkeivekakgaevvaidlesdtkdpmrgkivgvsldcfnppkayyfpfrhegleaqkqlpweafthlas  
liedpsvkkighnikydliilarygvtlkglegdtmlasyllldptrrthgldelaeevlghtmisykevtkelakesfarvplekakvyaced  
ahvtyllyqyfwpkkeeslwkvfteidrplievlahmemvgikidtaylrglsremaeklkeleekiytlagekfninsskqlgqilfeklkl  
ptvkktpkktaystdnevleelsavhelprlileyrtlaklkstyvdalpkmvnpetgrlhtsfngtvtatgrlsssdpnlnipvrgeegлки  
rqafvpeegflflsadysqidrlvlahysgdetlikafwggedihrrtaaeifgippeevtpemrrmaktinfgivygisdyglaqnlnitrke  
aaefieryfasfpgvkqymenivqeakqkgyvtllhrrrylpditsrnfvnrsfaertamntpiqgsadiiklamikihrifkekgfgtrml  
lqvhdelifepkeieeiqpivrqimegvvelkvplkvnlaigknwaeaka

### TG\_C3:

atggcctcttaaaggaacttcagctactaaaaccctttcgatgaccagatacagctggttcttgaccggataaagtaaaagaaattgta  
gaaaaggccaaagggccgaagtgggtgctattgacctgaaagtgtacgaaagaccccatgctgggaaaatagtaggggtctcgctttggt  
ttaaccggcccaaagcctattatttcccttttagacatgaaggccttgaggcccaaaagcagcttccctgggaggcctttactcatctggccag  
cctcattgaagacccctcagttaaaaagataggccacaatatcaagtatgacttgattattcttgctcgctacggcgtaactttaagggcctt  
gaaggggataccatgctggtctcgatctccttgatccaacacgctcgtaaccacggccttgatgagctggccgaagaggtcctggggcatacca  
tgatttcttacaaggaagtgtactaaagaactggccaaaggagagagctttgccagggtccctcttgaaaaggcaaaagtttacgcctgtgaaga  
cgccacgttacctatctgctttatcaatatttctggcccaaactcaaaggagaaagcctctggaaggtctttacggaattgatcgaccttta  
atagaagttttggcccatggaatggtaggtattaagattgacaccgcctatcttagaggactttcgcgagaaatggctgaaaagttaaagg  
agcttgaagaaaaatttacacctggctggtgaaaaatttaatatcaattccagcaacaactgggccagattttatttgaaaagctaaaact  
ccttacggttaaaaagacccccaaaaaaacggcctattcaacggataacgaagtattagaggaaactttctgcggtccacgaacttccgcgtctg  
atacttgagtatagaactctggctaaaactcaaactctacttatgttgatgcccctccgaagatggttaactctgaaactggctgtcttcatactt  
cctttaaccagacggttacggccactggaagactttcaagcagtgaccctaattcttcaaaatattcctgtgcgtggtgaagaggggcttaagat  
tcgccaggcctttgtgcggaggagggtttctttttctctcagctgactattctcagattgaattgcgcgctcctcgcccatatcgccgatga  
cgacaatctaattgaagcgttccaacgcgatttgatattcacacaaaaacggcgatggacattttccatgtgagcgaagaggaagtacggcc  
aacatgcgccgcaggcaaaaggccgttaacttcggtatcgtttacggatttagcgattacggattggcgcaaaacttgaacattacgcgcaagg  
aagctgccaatattatcgaaacttacttcgccagctttccggcgtaaaagcagtatattggaacacattgtgcaagaagcgaacagaaaggata  
tgtgacaacgctgttgcatcgccgcgctatttgctgatattacaagccgaatttcaacgtccgcagttttgcagagcggacggccaatgaa  
cacgcccattcaaggaagcgcgctgatattatcaagctcgcgatgataaaaaattcaccggatttttaagaaaaaggcctttgggacaaggat  
gcttcttcagggtgatgacgagcttatttttgaagcgccaaaagagattgaagaaatccagccaattgtccgacaaatcatggaaggagtggt  
gaattgaagggttcctctaaaagtaaacctggcaatagggaataattgggcagaggcaaaggcaataa

Translated protein sequence:

mgllkelpatktlsmtryelvldpdkvkeivekakgaevvaidlesdtkdpmrgkivgvsldcfnppkayyfpfrhegleaqkqlpweafthlas  
liedpsvkkighnikydliilarygvtlkglegdtmlasyllldptrrthgldelaeevlghtmisykevtkelakesfarvplekakvyaced  
ahvtyllyqyfwpkkeeslwkvfteidrplievlahmemvgikidtaylrglsremaeklkeleekiytlagekfninsskqlgqilfeklkl  
ptvkktpkktaystdnevleelsavhelprlileyrtlaklkstyvdalpkmvnpetgrlhtsfngtvtatgrlsssdpnlnipvrgeegлки  
rqafvpeegflflsadysqelrvlahiaddnlieafqrdldihtktamdfhvseevtanmrrqakavnfgivygisdyglaqnlnitrke  
aaefieryfasfpgvkqymenivqeakqkgyvtllhrrrylpditsrnfvnrsfaertamntpiqgsadiiklamikihrifkekgfgtrml  
lqvhdelifepkeieeiqpivrqimegvvelkvplkvnlaigknwaeaka

## ii. Tin-Taq Pol I

### **TinTaq Chimera1 (N-terminal) (C1):**

TinTatC1\_U            5'-cttgaggcccaaaagcagcttcccataaaagccctcaggacac-3'  
TinTatC1\_L            5'-aggtccctgagggcctttataggaagctgcttttgggcctcaag-3'

gllkelpatktlsdryelvldpdkvkeivekakgaevaidlesdtkdpmrgkivgvsicfnppkayyfpfrhegleaqqlkykalrdlke  
argllakdlsvlalreglglppgddpmlaylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavl  
ahmeatgvrl dvaylralslevaeearleaevfrlaghpfnlnsrldqlervlfdelglpaigktektgkrstsaavlealreahpivekilqy  
reltklkstyidplpdlihpgrlhrfnqtatatgrlsssdpnlnipvrtpglgrrirrafiaegwllvaldysqielrvlahlsgdenli  
rvfqegr dihtetaswmgfvpreadvplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetl  
fgrrryvpdlearvksvreaaermafnpvggtaadlmklamvklfprleemgarmllqvhdelvleapker aeavarlakevmegvyplavpl  
evevgigedwlsake

### **TinTaq Chimera2 (N-terminal back loop) (C2):**

TinTatC2\_U            5'-tcgtgttcaccgtgcgccggaaccctgggaggcctttactcatct-3'  
TinTatC2\_L            5'-agatgagtaaaggcctcccagggttccggcgcacggtgaacacga-3'  
TinTatC2\_U2           5'-tggcttcgtatctccttgatccatccaacaccacccccgagggggt-3'  
TinTatC2\_L2           5'-accccctcgggggtggtgttggatggatcaaggagatacgaagcca-3'

gllhefgllespkaleeapwpppegafvgfvlrskep mwadllalaaaggrvrhrapeweafthlasliedpsvkkighnikydlilaryg  
vtlkglegdtmlasyllcsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrl dvaylr  
alslevaeearleaevfrlaghpfnlnsrldqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpd  
lihpgrlhrfnqtatatgrlsssdpnlnipvrtpglgrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegr dihtetasw  
mgfvpreadvplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetlfgrrryvpdlearvks  
vreaaermafnpvggtaadlmklamvklfprleemgarmllqvhdelvleapker aeavarlakevmegvyplavplevevgigedwlsake

### **TinTaq Chimera3 (N-ter front loop) (C3):**

TinTatC3\_U            5'-tcctcgctacctcctggacccaacacgtcgtacccacggcct-3'  
TinTatC3\_L            5'-aggccgtgggtacgacgtgttgggtccaggaggtaggcgagga-3'  
TinTatC3\_U2           5'-acctttaatagaagttttggccacatggaggccacgggggtgc-3'  
TinTatC3\_L2           5'-gcacccccgtggcctccatgtgggcacaaaacttctattaaaggt-3'

mgllhefgllespkaleeapwpppegafvgfvlrskep mwadllalaaaggrvrhrapepykalrdlkeargllakdlsvlalreglglppgdd  
pmlaylldcrrthgldelaeevlghtmisykev tkelakgesfarvplekakvyacedahvtyllyqyfwpkkeeslwkvfteidrpliev  
lmeatgvrl dvaylralslevaeearleaevfrlaghpfnlnsrldqlervlfdelglpaigktektgkrstsaavlealreahpivekilq  
yreltklkstyidplpdlihpgrlhrfnqtatatgrlsssdpnlnipvrtpglgrrirrafiaegwllvaldysqielrvlahlsgdenl  
irvfqegr dihtetaswmgfvpreadvplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvet  
lfgrrryvpdlearvksvreaaermafnpvggtaadlmklamvklfprleemgarmllqvhdelvleapker aeavarlakevmegvyplavp  
levevgigedwlsake

#### **TinTaq Chimera4 (Outer thumb) (C4):**

TinTatC4\_U            5'-Agaggccccctttccgctgtcctgcccacatggaaatggtaggtat3'  
TinTatC4\_L            5'-Atacctaccattttccatgtgggccaggacagcggaaggggcctct3'  
TinTatC4\_U2           5'-Tgaagaaaaaatttacaccctggcggccaccccttcaacctcaact3'  
TinTatC4\_L2           5'-agttgaggttggaaggggtggccagccaggtgttaaattttttcttca3'

mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmemvgikidtaylrslremaek  
lkeleekiyltghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlihpgrtl  
htrfnqtatatgrlsssdpnlnipvrtplgqrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetlfgrrryvpdlearvksvreaaerma  
fnmpvqgtaadlmklamvklfprleemgarmllqvhdelvleapkeravearlakevmegvyplavplevevgigedwlsake

#### **TinTaq Chimera5 (Tip of thumb) (C5):**

TinTatC5\_U            5'-gaggccgaggtcttccgcctggcgtggtagaaaaatttaatatcaat-3'  
TinTatC5\_L            5'-attgatattaaatttttcaccagccaggcggaagacctcggcctc-3'  
TinTatC5\_U2           5'-agaggaactttctgcggtcacacccatcgtagagaagatcctgca-3'  
TinTatC5\_L2           5'-tgcaggatcttctccacgatgggggtggaccgcagaaagttcctct-3'

mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrladvaylralslevae  
iarleaevfrlgkfninsskqlgqilfeklklptvkktpkktaystdnevleelsavpivekilqyreltklkstyidplpdlihpgrtl  
htrfnqtatatgrlsssdpnlnipvrtplgqrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetlfgrrryvpdlearvksvreaaerma  
fnmpvqgtaadlmklamvklfprleemgarmllqvhdelvleapkeravearlakevmegvyplavplevevgigedwlsake

#### **TinTaq Chimera6 (Inner thumb) (C6):**

TinTatC6\_U            5'-tcctggaggccctccgcgaggccacgaacttccgcgtctgatact3'  
TinTatC6\_L            5'-agtatcagacgcggaagttcgtgggcctcgcgaggggcctccagga3'  
TinTatC6\_U2           5'-tgccctcccgaagatgggtaatcctaggacgggccctccacacccgct3'  
TinTatC6\_L2           5'-agcgggtgtggaggcggcccgctcctaggattaaccatcttcgggagggca3'

mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrladvaylralslevae  
iarleaevfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahelprrlileyrtilaklkstyvdalpkmvnrtgrtl  
htrfnqtatatgrlsssdpnlnipvrtplgqrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetlfgrrryvpdlearvksvreaaerma  
fnmpvqgtaadlmklamvklfprleemgarmllqvhdelvleapkeravearlakevmegvyplavplevevgigedwlsake

### **TinTaq Chimera7 (palm) (C7):**

TinTatC7\_U 5'tgaccaccttgccggacctcatccacccctgaaactggctcgtcttcatact3'  
TinTatC7\_L 5'agtatgaagacgaccagtttcagggtggatgaggtccggcaaggggtca3'  
TinTatC7\_U2 5'tgaccctaattcttcaaaatattccctgtccgcacccccgcttgggcaga3'  
TinTatC7\_L2 5'tctgccaagcgggggtgcggacaggaatattttgaagattaggggtca3'

mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvlalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrl dvaylralslevae  
iarleaevfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlihsetgrl  
htsfngtvtatgrlsssdpnlniivrtplgqrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetlfgrrryvpdlearvksvreaaerma  
fnmpvggtaadlmklamvklfprleemgarmllqvhdelvleapkeraeavarlakevmegvyplavplevevgigedwlsake

### **TinTaq Chimera8 (Back of palm to thumb) (C8):**

TinTatC8\_U 5'Tccgatcccaacctccagaacatccctgtgcgtgggtgaagaggggct3'  
TinTatC8\_L 5'Agccccctcttcaccacgcacagggatgttctggaggttgggatcgga3'  
TinTatC8\_U2 5'Attgatctgcgagtttttagcccatctctccggcgacgagaacctgat3'  
TinTatC8\_L2 5'atcaggttctcgtcgccggagagatgggctaaaactcgcatatcaat3'

mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvlalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrl dvaylralslevae  
iarleaevfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlihprtgrl  
htrfnqtatatgrlsssdpnlniivrggeglkirqafvpeegflflsadysqidlrvlaflsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetlfgrrryvpdlearvksvreaaerma  
fnmpvggtaadlmklamvklfprleemgarmllqvhdelvleapkeraeavarlakevmegvyplavplevevgigedwlsake

### **TinTaq Chimera9 (Back of fingers) (C9):**

TinTatC9\_U 5'-Agatagagctcagggtgctggcccatctactcgggagatgaaacctt3'  
TinTatC9\_L 5'-Aaggtttcatctcccagtaatgggccagcaccctgagctctatct3'  
TinTatC9\_U2 5'-Tgcggcggtatggccaagactataaacttcggggtcctctacggcat3'  
TinTatC9\_L2 5'-atgccgtagaggacccccgaagtttatagtcttggccatacgccgca3'

mgllhefgllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvlalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrl dvaylralslevae  
iarleaevfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlihprtgrl  
htrfnqtatatgrlsssdpnlnipvrtplgqrrirrafiaegwllvaldysqielrvlaflsgdetlikafwqgedihrrtaaeifgippeev  
tpemrrmaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvrawiektleegrrrgyvetlfgrrryvpdlearvksvreaaerma  
fnmpvggtaadlmklamvklfprleemgarmllqvhdelvleapkeraeavarlakevmegvyplavplevevgigedwlsake



### **TinTaq Chimera10 (Tip of fingers) (C10):**

TinTatC10\_U      5'-tgatgcgccggggcgccaagaccatataaactttggcattgtttacggca-3'  
TinTatC10\_L      5'-Tgccgtaaacaatgccaaagtttatgggtcttggccgcccggcgcatca-3'  
TinTatC10\_U2     5'-Tactgtggagacccttttcggacggcgcgcgtacgtgccagacctaga-3'  
TinTatC10\_L2     5'-tctaggtctggcacgtagcggcgcgctccgaaaaggtctccacgta-3'

mgllehglllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrlldvaylralslevae  
iarleaevfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlihpgrl  
htrfnqtatatgrlsssdnpnlqnipvrtplgqrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgivygmspyglakelkigrreakafieryferypgvkrymeqivaearekgyvetlfgkrryvpdlearvksvreaaerma  
fnmpvgqtaadlmklamvklfprleemgarmllqvhdelvleapkeraeavarlakevmegvyplavplevevgigedwlsake

### **TinTaq Chimera11 (Front of fingers to C-ter) (C11):**

TinTatC11\_U      5'-Gtacgtggagaccctcttcggccggcaaaaggcctcttcctgaca-3'  
TinTatC11\_L      5'-Tgtcaggaagaggccttttgcggccgaagagggctctccacgtac-3'  
TinTatC11\_U2     5'-Gatattatcaagctcgccatgatagtgaagctcttccccaggctg-3'  
TinTatC11\_L2     5'-cagcctggggaagagcttcactatcatggcgagcttgataaatatc-3'

mgllehglllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrlldvaylralslevae  
iarleaevfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlihpgrl  
htrfnqtatatgrlsssdnpnlqnipvrtplgqrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvwrawiektleegrrrgyvetlfgkrrplpdinspnrtarefaerta  
intpiggtaadiiklamvklfprleemgarmllqvhdelvleapkeraeavarlakevmegvyplavplevevgigedwlsake

### **TinTaq Chimera12 (C-ter back of palm) (C12):**

TinTatC12\_U      5'-Gccgacctcatgaagctggctataaaaaattcaccggatttttaaga-3'  
TinTatC12\_L      5'-Tcttttaaaaatccggtgaatttttatagccagcttcatgaggtcggc-3'

mgllehglllespkaleeapwpppegafvgfvlsrkepmwadllalaaarggrvhrapepykalrdlkeargllakdlsvalalreglglppgdd  
pmllylldpsnttpegvarryggewteeageraalserlfanlwgrlegeerllwlyreverplsavlahmeatgvrlldvaylralslevae  
iarleaevfrlaghpfnlnsrdqlervlfdelglpaigktektgkrstsaavlealreahpivekilqyreltklkstyidplpdlihpgrl  
htrfnqtatatgrlsssdnpnlqnipvrtplgqrrirrafiaegwllvaldysqielrvlahlsgdenlirvfqegrdihtetaswmfgvpreav  
dplmrraaktinfgvlygmsahrlsqelaipyeeaqafieryfqsfpkvwrawiektleegrrrgyvetlfgrrryvpdlearvksvreaaerma  
fnmpvgqtaadlmklamvikihrifkekgfgrmlqvhdelfeapkeieeiqpivrqimegvvelkvplkvnlaignkwaeaka

#### iv. DNA polymerase I fusion sequences

***P.isCren7* ssDNA binding protein** (*E.co* codon optimised),  
GeneSys Biotech Ltd.

```
atg gaggaagtgctggatcgtgaatatgaagtagaatatggagggcgtaaataaccgccttaagcctgttaaagcatgggttctccagccccct  
ggcaaaccaggtgtcgtcatagccctctttaaactcccagacggcaagacaattcgaaaagtcattatgaaactgccgccaagt tag
```

Translation of DNA sequence:

```
1 atg gaggaagtgctggatcgtgaatatgaagtagaatatggagggcgtaaataaccgcctt  
1 M E E V L D R E Y E V E Y G G R K Y R L  
61 aagcctgttaaagcatgggttctccagccccctggcaaaccaggtgtcgtcatagccctc  
21 K P V K A W V L Q P P G K P G V V I A L  
121 tttaaactcccagacggcaagacaattcgaaaagtcattatgaaactgccgccaagt tag  
41 F K L P D G K T I R K V I M K L P P S *
```

#### N-terminal fusion (*P.isCren7*-*G.me\** Pol I)

Translation of DNA sequence:

```
meevldreyevygrkyrlkpvkawvlqppgkpgvvialfklpdgktirkvimklpps gth megekpleemefaivdviteemladkaalvve  
vmeenyhdapivgialvnehgrffmrpetaladsqflawladetkkksmfdakravvalkwgielrgvafdlllaayllnpaquadagdiaavak  
mkqyeavrsdeavygkgvkrslpdeqtlaelvrkaaiwaleqpfmddlrnneqdqltkleqplaaailaemeftgvnvtdtkrleqmgse  
qlraieqriyelagqefninspqlgvilfeklqlpvltkktgystadvleklaphheivenilhyrqlgklqstyieglkvvrpdtgkv  
tmfnqaltqtgrlssaepnlqnipirleegrkirqafvpsepdwlifaadysqielrvlahiaddnlieafqrdldihtktamdifhvseeev  
tanmrrqakavnfgivygisdyglaqnlitrkeaaefieryfasfpgvkqymenivqeakqgyvtllhrrrylpditsrnfvrfsaerta  
mntpiqgsaadiikkamidlaarlkeeqlqarlllqvhdelileapkeeierlcelvpevmeqavtlrvplkvdyhygptwydak
```

#### C-terminal fusion (*G.me\** Pol I-*P.isCren7*)

Translation of DNA sequence:

```
megekpleemefaivdviteemladkaalvvevmeenyhdapivgialvnehgrffmrpetaladsqflawladetkkksmfdakravvalkw  
gielrgvafdlllaayllnpaquadagdiaavakmkqyeavrsdeavygkgvkrslpdeqtlaelvrkaaiwaleqpfmddlrnneqdqltkl  
eqplaaailaemeftgvnvtdtkrleqmgseqlraieqriyelagqefninspqlgvilfeklqlpvltkktgystadvleklaphheiv  
enilhyrqlgklqstyieglkvvrpdtgkvhtmfntqaltqtgrlssaepnlqnipirleegrkirqafvpsepdwlifaadysqielrvlahi  
addnlieafqrdldihtktamdifhvseeevtanmrrqakavnfgivygisdyglaqnlitrkeaaefieryfasfpgvkqymenivqeakq  
gyvtllhrrrylpditsrnfvrfsaertamntpiqgsaadiikkamidlaarlkeeqlqarlllqvhdelileapkeeierlcelvpevme  
qavtlrvplkvdyhygptwydak gth meevldreyevygrkyrlkpvkawvlqppgkpgvvialfklpdgktirkvimklpps
```

## N-terminal fusion (*P.isCren7-T.in Pol I*)

Translation of DNA sequence:

meevldreyevyggrrkyrlkpvkawvlqppgkpgvvialfklpdgktirkvimklpps<sup>gth</sup>klklkelpatktlsydqyelvldpdkvkeive  
kakgaevvaialasdtkdpmrgkivgvsfcfnppkayyfpfrhegleaqqlpweafthlasliedpsvkkighnikydliilarygvtklgle  
gdtmlasyllldptrrthgldelaeevlghtmisykevtkelakgesfarvplekakvyacedahvtyllyqyfwpkkkeeslwkvfteidrpli  
evlahmemvgikidtaylrglsremaekleleekiytlagekfninsskqlgqilfeklklptvkktpkktaystdnevleelsavhelpri  
leyrtlaklkstyvdalpkmvnmpetgrlhtsfngtvtatgrlsssdpnlnipvrgeeglkirqafvpeegflflsadysqidrlvlahysgde  
tlkafwqgedihrrtaaeifgippeevtpemrrmaktinfgivymspyglakelkigrreakafieryferypgvkrymeqivaearekgyv  
etlfgkrplpdinspnrtarefaertaintpiqgtaadiiklamikihrifkekgfgtrmlqvhdelldfevpekeieeiqpivrqimegvve  
lkvplkvnlaigknwaeaka

## C-terminal fusion (*P.isCren7-T.in Pol I*)

Translation of DNA sequence:

klklkelpatktlsydqyelvldpdkvkeivekakgaevvaialasdtkdpmrgkivgvsfcfnppkayyfpfrhegleaqqlpweafthlas  
liedpsvkkighnikydliilarygvtklglegdtmlasyllldptrrthgldelaeevlghtmisykevtkelakgesfarvplekakvyaced  
ahvtyllyqyfwpkkkeeslwkvfteidrplievlahmemvgikidtaylrglsremaekleleekiytlagekfninsskqlgqilfeklkl  
ptvkktpkktaystdnevleelsavhelprileyrtlaklkstyvdalpkmvnmpetgrlhtsfngtvtatgrlsssdpnlnipvrgeeglk  
rqafvpeegflflsadysqidrlvlahysgdetlikafwqgedihrrtaaeifgippeevtpemrrmaktinfgivymspyglakelkigrre  
akafieryferypgvkrymeqivaearekgyvetlfgkrplpdinspnrtarefaertaintpiqgtaadiiklamikihrifkekgfgtrml  
lqvhdelldfevpekeieeiqpivrqimegvvelkvplkvnlaigknwaeakagth<sup>meevldreyevyggrrkyrlkpvkawvlqppgkpgvvialfklpdgktirkvimklpps</sup>

## v. Plasmids

### i. Design of pET24a+His vector

The pET24a(+) vector (Novagen, USA) was modified to add a 6x His tag upstream of the *Nde I* site. The His tag was inserted between *XbaI* and *BamHI* sites using the primer:

New 6xHis-tagged region:

ttccctcttagaataattttgtttaactttaagaaggagatatactatgcaccatcaccatcaccat<sup>atg</sup>gcta  
gcggatccgaattc

## vi. Set for Britain Award 2013

Work from this thesis was entered into 'Set for Britain 2013' a national competition organised by the UK Parliamentary and Scientific Committee for early-stage researchers. My poster and presentation at the Houses of Parliament to Members of Parliament and selected dignitaries during National Science and Engineering Week was awarded 2<sup>nd</sup> Prize from over 500 entries ([www.setforbritain.org.uk](http://www.setforbritain.org.uk))

